



(12) 发明专利

(10) 授权公告号 CN 113383345 B

(45) 授权公告日 2024.10.18

(21) 申请号 202080004735.5

(22) 申请日 2020.12.17

(65) 同一申请的已公布的文献号
申请公布号 CN 113383345 A

(43) 申请公布日 2021.09.10

(30) 优先权数据
16/718,071 2019.12.17 US

(85) PCT国际申请进入国家阶段日
2021.02.25

(86) PCT国际申请的申请数据
PCT/US2020/065680 2020.12.17

(87) PCT国际申请的公布数据
W02021/127225 EN 2021.06.24

(73) 专利权人 索尼互动娱乐有限责任公司
地址 美国加利福尼亚州

(72) 发明人 A.加卢坦

(74) 专利代理机构 北京市柳沈律师事务所
11105

专利代理师 张晓明

(51) Int.Cl.
G06N 3/006 (2023.01)
G06N 3/0464 (2023.01)
G06N 5/02 (2023.01)
G06F 18/241 (2023.01)
G06N 3/045 (2023.01)
G06N 3/094 (2023.01)
G06N 3/0475 (2023.01)
G06N 3/044 (2023.01)

(56) 对比文件
US 2012284080 A1, 2012.11.08
US 2012290511 A1, 2012.11.15
US 2017160813 A1, 2017.06.08

审查员 陈萌

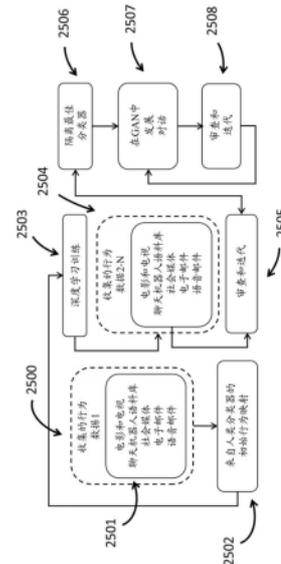
权利要求书2页 说明书25页 附图32页

(54) 发明名称

用于定义情感机器人的方法和系统

(57) 摘要

公开了一种用于训练智能代理的方法,其包括:形成人格矩阵;将认知偏见矩阵与所述人格矩阵组合;以及基于所组合的认知偏见矩阵和人格矩阵针对情景生成行为函数。



1. 一种用于训练智能代理的方法,其包括:

- a) 形成人格矩阵;
- b) 将认知偏见矩阵与所述人格矩阵组合;
- c) 基于所组合的认知偏见矩阵和人格矩阵针对情景生成行为函数,

其中,所述情景包括情景基线和用于分析认知偏见的行为掩码,所述情景基线包括基本人格元素、表示成长环境对行为的影响的发展过滤器和表示与人和环境的长期关系对行为的影响的关系过滤器,所述基本人格元素通过所述发展过滤器修改、通过所述关系过滤器情境化并通过所述行为掩码进一步修改,以创建所述行为函数的基本作用背景。

2. 如权利要求1所述的方法,其中所述人格矩阵至少包括对应于迈尔斯布里格斯类型指标的参数。

3. 如权利要求1所述的方法,其中所述人格矩阵包括对应于大五人格特质的参数。

4. 如权利要求1所述的方法,其中所述人格矩阵至少包括一个或多个传记参数。

5. 如权利要求4所述的方法,其中所述一个或多个传记参数包括位置、历史、文化或教育参数。

6. 如权利要求1所述的方法,其中所述行为函数根据分类人际交互数据进行修改,并且其中所述人际交互数据基于所述人格矩阵的参数进行分类。

7. 如权利要求6所述的方法,其中所述行为函数由基于所述分类人际交互数据训练的神经网络生成,并且其中所述神经网络训练由所述人格矩阵修改。

8. 一种用于使用智能代理的方法,其包括:

- a) 确定情景;
- b) 基于所组合的认知偏见矩阵和人格矩阵针对所述情景应用行为函数以生成对所述情景的响应,

其中,所述情景包括情景基线和用于分析认知偏见的行为掩码,所述情景基线包括基本人格元素、表示成长环境对行为的影响的发展过滤器和表示与人和环境的长期关系对行为的影响的关系过滤器,所述基本人格元素通过所述发展过滤器修改、通过所述关系过滤器情境化并通过所述行为掩码进一步修改,以创建所述行为函数的基本作用背景。

9. 如权利要求8所述的方法,其中确定所述情景包括至少使用利用自然语言处理数据训练的神经网络。

10. 如权利要求9所述的方法,其中所述自然语言处理数据包括对用户问题的至少一个预先回答的响应。

11. 如权利要求9所述的方法,其中所述行为函数修改由所述神经网络确定的响应。

12. 如权利要求8所述的方法,其中所述人格矩阵和所述认知偏见矩阵至少包括幅值和权重。

13. 如权利要求8所述的方法,其还包括:c) 监视所述响应的成功指示。

14. 如权利要求13所述的方法,其中监视成功指示包括使用机器视觉或语音分析来确定情感观点。

15. 如权利要求13所述的方法,其中监视成功指示包括监视对用户的响应延迟、回调至同一主题、对所述用户的不屑一顾或对所述用户的愤怒。

16. 如权利要求13所述的方法,其还包括:使用响应的所述成功指示来调整所述认知偏

见矩阵或所述人格矩阵。

17. 如权利要求8所述的方法,其中确定所述情景包括使用来自机器视觉或对象识别的视觉信息。

18. 如权利要求8所述的方法,其中所述行为函数由基于分类人际交互数据训练的神经网络应用,并且其中所述神经网络训练由所述人格矩阵修改。

19. 一种智能代理系统,其包括:

处理器;

联接到所述处理器的存储器;

嵌入在所述存储器中的非暂时性指令,所述非暂时性指令在被执行时致使所述处理器执行包括以下的方法:

a) 确定情景;

b) 基于所组合的认知偏见矩阵和人格矩阵针对所述情景应用行为函数以生成对所述情景的响应,

其中,所述情景包括情景基线和用于分析认知偏见的行为掩码,所述情景基线包括基本人格元素、表示成长环境对行为的影响的发展过滤器和表示与人和环境的长期关系对行为的影响的关系过滤器,所述基本人格元素通过所述发展过滤器修改、通过所述关系过滤器情境化并通过所述行为掩码进一步修改,以创建所述行为函数的基本作用背景。

20. 一种计算机可读介质,包括非暂时性指令,所述非暂时性指令在被执行时致使计算机实现包括以下的方法:

a) 确定情景;

b) 基于所组合的认知偏见矩阵和人格矩阵针对所述情景应用行为函数以生成对所述情景的响应,

其中,所述情景包括情景基线和用于分析认知偏见的行为掩码,所述情景基线包括基本人格元素、表示成长环境对行为的影响的发展过滤器和表示与人和环境的长期关系对行为的影响的关系过滤器,所述基本人格元素通过所述发展过滤器修改、通过所述关系过滤器情境化并通过所述行为掩码进一步修改,以创建所述行为函数的基本作用背景。

用于定义情感机器人的方法和系统

技术领域

[0001] 本公开的方面涉及专家系统,具体地,本公开的方面涉及使用心理学和社会学信息开发专家系统和机器学习以用于进行更大的行为复制。

[0002] 发明背景

[0003] 智能系统(意指一起工作的机器和网络服务)与人类相比可具有更大的捕获、记忆和比较听觉、视觉和其他感官提示的能力。

[0004] 从视觉开始,摄像技术已改进到一定水平,在所述水平下,智能系统(IS)可看到人类能够看到的一切并且可看到更多一后方、上方、下方、长距离处、在几乎无光的情况下以及在人眼不可见但一些动物能够探测到的频率范围(像红外光和紫外光)内。事实上,IS可看到其他电磁波,像X射线、微波和无线电波。为了理解人类将看到什么,IS将了解人类视觉的局限性,并且考虑如果是人类他们会看到什么,但其他数据也可用(如果需要的话),从而形成超人的视觉(以及所见事物的超人的记忆力和准确度)。

[0005] 至于听觉,高分辨率双声道录音已经非常接近地模仿了人类听到的内容,但为何停滞不前了?正在开发能够将个体声音或符尾与演出分开的系统,所述系统将使IS能够对听觉环境的个体元素进行观察。

[0006] 味觉和嗅觉呢?“自1982年以来,一直在进行研究以开发通常称为电子鼻的技术,所述技术可检测和识别气味和风味。”从那时起,随着对嗅觉识别工作方式的更好理解,工作取得了显著进展,其中来自麻省理工学院的Andreas Mershin和Shuguang Zhang凭借其纳米鼻(Nano-Nose)获得DARPA奖。

[0007] 通过触摸识别对象的能力是人类感觉的核心功能之一。BioTac触觉传感器具有“利用性能以及人类感知力(并且有时,性能甚至优于人类感知力)以基于对象的柔顺性、质地和热特性来辨别和识别对象的能力。”

[0008] 来自斯坦福学院的Yilun Wang、Michal Kosinski在2017年进行了一项研究,所述研究证明深度神经网络能够以比人类对照组更高的准确度根据面部图像检测性取向。

[0009] 现在,一些系统能够比人类更准确地解读情感和行为提示。来自麻省理工学院的情感计算实验室的科学家已经开始推销“人工情感智能”,所述“人工情感智能”是从Paul Ekman和Wallace V Friesen在20世纪80年代开发但一直更进一步发展的情感面部动作编码系统(EMFACS)演变而来。

[0010] 本公开的方面正是在此背景下产生的。

附图说明

[0011] 通过以下具体实施方式结合附图可容易理解本发明的教导内容,其中:

[0012] 图1是根据本公开的方面的IS的组分的图解概述。

[0013] 图2是根据本公开的方面的作为一系列层的人类行为的视图。

[0014] 图3是描绘根据本公开的方面的迈尔斯布里格斯人格分型连续统一体的框图。

[0015] 图4是根据本公开的方面的大五人格特质的视图。

- [0016] 图5是描绘根据本公开的方面的情景基线以及感觉输入和输出的框图。
- [0017] 图6是根据本公开的方面的发展过滤器的参数的图解视图。
- [0018] 图7是根据本公开的方面的关系过滤器的元素的说明图。
- [0019] 图8是根据本公开的方面的行为掩码的元素的说明图。
- [0020] 图9是根据本公开的方面的包括行为函数的心智堆栈的说明图。
- [0021] 图10是描绘根据本公开的方面的行为函数的详述元素的图。
- [0022] 图11是示出根据本公开的方面的从DNA一直到基于行为的观察函数的完整堆栈的框图。
- [0023] 图12是示出根据本公开的方面的基线人格面貌与智能系统的单个实例之间的连接的图。
- [0024] 图13是描绘根据本公开的方面的行为偏见到各种过滤器和掩码的映射的框图。
- [0025] 图14是示出根据本公开的方面的专家系统如何用于映射行为或认知偏见的图。
- [0026] 图15是描绘根据本公开的方面的构成人格基线的层的框图。
- [0027] 图16是示出根据本公开的方面的MBTI权重的表。
- [0028] 图17是描绘根据本公开的方面的文化层的框图。
- [0029] 图18是示出根据本公开的方面的行为集合的框图。
- [0030] 图19是描绘根据本公开的方面的如何针对行为偏见映射情景基线以形成情景偏见集的框图。
- [0031] 图20是示出根据本公开的方面的将心理学参数映射到行为偏见从而向每个行为偏见赋予权重的框图。
- [0032] 图21是示出根据本公开的方面的IS实例到情景的行为偏见矩阵组成的示例的表。
- [0033] 图22是描绘根据本公开的方面的将偏见归咎于每个IS的情景环境的框图。
- [0034] 图23是示出根据本公开的方面的人格矩阵的示例的表。
- [0035] 图24是示出根据本公开的方面的基线人格矩阵的替代视图的表。
- [0036] 图25是描绘根据本公开的方面的捕获和分析行为数据的框图。
- [0037] 图26是示出根据本公开的方面的社会分类的视图的框图。
- [0038] 图27是根据本公开的方面的包括完整人格矩阵的参数的表视图。
- [0039] 图28是根据本公开的方面的用于描述人格的矩阵的未标记矩阵视图。
- [0040] 图29A是根据本公开的方面的用于在智能系统中使用的递归神经网络的简化节点图。
- [0041] 图29B是根据本公开的方面的用于在智能系统中使用的展开的递归神经网络的简化节点图。
- [0042] 图29C是根据本公开的方面的用于在智能系统中使用的卷积神经网络的简化图。
- [0043] 图29D是根据本公开的方面的用于在智能系统的开发中训练神经网络的方法的框图。
- [0044] 图30是描绘根据本公开的方面的在智能系统中训练生成对抗神经网络的框图。
- [0045] 图31描绘根据本公开的方面的智能代理系统的框图。

具体实施方式

[0046] 尽管以下具体实施方式包含用于说明目的的许多具体细节,但本领域的任何普通技术人员应了解,对以下细节的许多变化和更改在本发明的范围内。因此,以下所述的本发明的示范性实施方案在不失一般性并且未强加对要求保护的本发明的限制的情况下进行阐述。

[0047] 引言

[0048] 机器尚未精通人际交互的技术。尽管许多聊天机器人曾偶尔愚弄过用户,但与计算机的通信通常是重复逻辑的并且明显是非人的。人类通常不会理性地采取行动。我们有许多认知偏见。然而,这些行为是“可预见地非理性的”。没有任何事物能阻止智能机器以与人类采取行动的方式相同的方式来“非理性地”采取行动。

[0049] 在行为经济学领域,已在有限理性、前景理论、跨期选择、助推理论、行为金融和行为博弈理论方面进行了大量研究。随着这些理论发展并且变得更具决定性,没有任何事物能阻止智能机器以与人类在做出经济决策时将采取行动的方式相同的方式来采取行动。替代地,没有任何事物能阻止经IS增强的人类变得不太理性。

[0050] 本公开的方面可分解成多个组分,所述多个组分在合在一起时将提供用于定义和构建智能系统或IS的完整系统。这种IS包括以下中的一些或全部:装置、网络、存储设备、数据结构、处理、算法、输入、输出和各种人工智能技术,包括但不一定限于深度神经网络、卷积神经网络、重组神经网络、专家系统生成对抗网络和使用训练和/或推断的人工神经网络。目的是教导IS—包括从简单的聊天机器人到完整的类人机器人的实例—以更像人类的方式采取行动。为此,我们将评估人类的心理学、社会学和物理学响应以及智能系统(IS)能够模仿他们的方式。

[0051] 所述问题可分解成多个组分。着眼于图1,IS100可记录101与人类具有的输入相同的输入:视觉102、听觉103、触觉104、味觉105和嗅觉106。接下来,IS可(最终几乎实时地)分析这些输入并且计算108以及使用触感、语音生成和机器人技术执行响应109。由IS构建的响应应能够以其他人类无法将其与人类区分的方式模仿人类响应并且潜在地甚至更加同理的(或马基雅维里的(Machiavellian))。本公开的大部分将解决这些理解和响应的社会和心理方面。最终,这些系统能够如何用于生成虚拟环境并且与现实世界产生反应并在现实世界中进行反应?让我们稍微更深入地探讨高层级体系结构。

[0052] 输入

[0053] 稍微更详细地着眼于输入,智能系统(意指一起工作的机器和网络服务)与人类相比可具有更大的捕获、记忆和比较听觉、视觉和其他感官提示的能力。

[0054] 视觉

[0055] 从视觉102开始,当然,摄像技术已改进到一定水平,在所述水平下,IS可看到人类能够看到的一切,并且在本公开的一些方面中可看到更多。后方、上方、下方、长距离处、在几乎无光的情况下以及在人眼不可见但一些动物能够探测到的频率范围(像红外光和紫外光)内。事实上,IS可看到其他电磁波(通过使用此类传感器),像X射线、微波和无线电波。为了理解人类将看到什么,IS将被编程以了解人类视觉的局限性,并且考虑如果他们是人类他们会看到什么,但其他数据也可用(如果需要的话),从而形成超人的视觉(以及所见事物的超人的记忆力和准确度)。使用机器视觉和/或对象识别,IS可能检测包括人类在内的

物理世界中的对象并且对其进行分类,从而制定实际的人类响应。

[0056] 听觉

[0057] 关于听觉103,高分辨率双声道录音已经非常接近地模仿了人类听到的内容和听的方式。根据本公开的方面,IS可将声音录制为对象,正如像杜比全景声的系统将音频播放为离散对象,来自像General Harmonics、Celemony、Red Pill和Sony的公司的其他系统也正在开发中,所述系统可从立体声或环绕声场捕获声音并且将其分成单个元素。例如且不限于,IS或经IS辅助的人类可听交响乐并且只想听第一小提琴或只想听法国号,并且IS可只隔离那些乐器,从而利用精细化控制(使用频谱分解方法,诸如单个分量分析(ICA))基本上将现实世界变成录音室。在一些实施方案中,IS可被整合到人类生物学中,并且人类在音乐会中可简单地想“我希望法国号更大声一点”,然后你(或你的半机械部件)可“改变混音”。通过此整合,你不但能够记住你所听到的一切,而且你能够以不同的混音回听所述内容。

[0058] 味觉和嗅觉

[0059] 电子鼻可识别人类会识别的气味和味道以及还有其他动物(例如狗)会识别的气味和味道。不同的人类对不同的气味和味道具有不同的响应,但了解我如何嗅闻与你如何嗅闻(用作动词)完全是机器学习数据分析问题。应注意,不涉及感受性质(Qualia),因为这只是识别问题。IS当然可学习鉴赏不同文化或不同个体所偏好的食物的味道和气味。事实上,如果IS要利用适当的硬件来吃(咀嚼和吞咽)食物,它也可知道口感(或更确切地将口感存储在其记忆力内)。IS将未必需要消化食物以获得能量(尽管其中可如此设计)。然而,机器可被训练来鉴赏(或更确切地由外人感知来鉴赏)作为人类或作为任何特定人类对食物和饮品的鉴赏的所有微妙之处。餐厅可雇用带有IS实例的Jonathan Gold机器来帮助评判他们的菜单。替代地,连锁餐厅可通过具有适当训练和硬件的IS厨师来基于他们的客户的喜好动态地调整调味料。

[0060] 触觉

[0061] 通过触摸识别对象的能力是人类感觉的核心功能之一。触摸传感器可用于检测初次接触、接触位置、滑动、曲率半径和边缘,以及确定三轴力以实现未知位置中具有未知特性的对象的灵巧触握。正如其他感觉一样,出于本公开的目的,将假设随着时间的流逝,高分辨率触摸(接收和给予)将变得更加可用。

[0062] 行为分析

[0063] 本公开的核心与我们生命体的心理学和情感方面有关。人类相当擅长识别其他人类的情感状态和癖性。一个人可辨别某人是生气还是快乐还是处于许多其他情感。分析他人的情感和倾向是基于人们所说的内容,但也基于对他们肢体语言、面部表情(包括微表情、声音的音色和音调、气味和物理观察(像潮红的皮肤、鸡皮疙瘩、眼泪等))的解读。一个人对此的诠释在一定程度上受他们自身的经验、成见和预期影响。偏执型的人认为别人在故意整他们,但是存在将一个人的预期映射到另一人的行为并且一般地映射到环境的许多更细微的型式。本公开将着眼于IS能够如何理解情感和行为倾向以及人类或进行得更好。

[0064] 根据本公开的方面,IS可考虑与所述IS交互的人的感受,例如,悲伤、喜悦、愤怒、同理心或调情。此外,根据本公开的方面的IS可与诸如电视、广告牌、展厅窗口中的销售员、价格变动或群体行为的其他输入交互并且从中不断学习。

[0065] IS必须能够解读并“理解”情感输入,然后才可对情感输入做出响应。本公开不专注于解读情感的原语。相反,本公开的方面采集所有元素并且分析它们对环境的心理学和社会学的含义,但将假设可使用其他技术来捕获基本原语。

[0066] 响应

[0067] 根据本公开的方面,IS可表现出对给定情感输入的适当响应。根据输入和输入周围的情况,IS可以情感(例如,同理心、愤怒、蔑视、群体思想)做出响应。在一些实现方式中,IS可利用行动(例如,购买、出售或决定采取行动(例如,清扫或烹调))来对输入和情况做出响应。一旦智能机器被教导来解读身体微提示(面部表情和其他肢体语言、嗅闻触摸等)中的情感,智能装置(例如,机器人)就可被教导来模仿诸如面部表情、气味、汗水等的身体行为,从而使得人类能够感受到他们的同理心、蔑视等。

[0068] 与IS对象一起进食。她/他可鉴赏食物的味道和气味并且适当地做出响应—甚至考虑到人的已知偏好。在吃完一顿难消化的膳食之后,她/他可能会表现出过于饱腹、无精打采、醺醉、性冲动。

[0069] 人类通常不会理性地采取行动。我们有许多认知偏见。然而,这些行为是“可预见地非理性的”。没有任何事物能阻止智能机器以与人类采取行动的方式相同的方式来“非理性地”采取行动。

[0070] 将剖析做出合理的、甚至非常合理且同理的(或者不太理想地,操纵的)响应所需的步骤。如图2所示,人类行为可被视为一系列层。存在对我们的感觉分析引擎200的多个输入。首先,存在如上所述的感觉记录201。还存在环境因素202,其随后将被更充分地描述。构建响应计算203所需的其他关键元素是一般地塑造我们的行为的人格元素204。行为掩码205—指代当我们向他人展示自己时戴着的面具,并且也指代计算机编程中的掩码,所述计算机编程指代过滤掉各种位或字节以将其从输出中移除。此后,将论述行为函数206,其使用数学意义上的函数,在所述函数中你具有对输入集进行运算以形成所需输出的算法或运算集,但将其应用于心理学决策。

[0071] 情景基线

[0072] 情景基线是IS引入任何情景或交互的基本人格结构。存在情景基线的三个组分:第一是基本人格类型或人格元素。在人类中,这些主要是遗传基因学和幼儿期的结果,并且通常表示人类的基本观点(例如受虐的儿童通常永远学不会信任)。从复制人类响应能力的角度来看,下一个发展层是发展过滤器。发展过滤器是我们基本人格之上的文化和社会覆盖。这由我们的社会和文化环境驱动,所述社会和文化环境可包括家庭、社区、朋友等。第三元素是关系过滤器。这些是基于背景作用于我们的过滤器。这反映与当前位置和人的预先存在的关系。

[0073] 如本文所用,基本人格元素指代基本人类特质的量化和分析。无疑存在这些特质的基因组分,并且将来基因组分无疑将在基本人格元素的分析中发挥作用。在一些实施方案中,基本人格元素局限于用于基本人格分析的心理学(和数据)方法。在替代实施方案中,甚至还考虑基因特质和体质,可使用基因图谱和由于基因密码标记而产生某些人格特质的可能性来应用此类基因特质。

[0074] 本领域中的任何人格塑造系统可用于基本人格元素。例如且不限于,在心理学文献中通常使用两种不同模型中的任一种来描述人格类型。一种模型是迈尔斯布里格斯人格

分型,其基于荣格原型并且将人格分成16种二元项组合,如图3所示:

[0075] -外向301->内向302

[0076] -实感303->直觉304

[0077] -思维305->情感306

[0078] -判断307->知觉308

[0079] 另一种常见的人格分析工具是大五人格特质或五因素模型,其最初基于Ernest Tupes和Raymond Christal的著作,并且随后由J.M.Digman和Lewis Goldberg改进,并且有人认为这表示所有人格特质背后的基本结构。

[0080] 如图4所示,大五人格特质400通常被描述为:

[0081] -经验开放性(有创造力的/好奇的与始终如一的/谨慎的) 401

[0082] -尽责性(高效的/有组织的与悠闲的/粗心的) 402

[0083] -外向性(开朗的/精力充沛的与独处的/高冷的) 403

[0084] -亲和性(友好的/富同情心的与挑战的/超然的) 404

[0085] -情绪不稳定性(敏感的/紧张的与无虑的/自信的) 405

[0086] 还存在众多其他人格连续统一体,包括由人格学家研究的以下特质。这些特质包括但不限于:

[0087] -马基雅维里主义:指代通常通过不诚实行为来操纵他人行为的个体。马基雅维里主义者通常对金钱和权力感兴趣,并且在此追求中务实地利用他人,

[0088] -成就感需要:高度需要成就感的那些人想要达成很多,并且为他们自己设定高卓越标准。他们能够持续且努力地为遥远的目标而工作。

[0089] -认知需要:高度需要认知的人们发现理解事物是有益的,并且愿意在此追求中耗费大量认知努力。此类个体喜欢学习以及尝试理解新事物的过程。

[0090] -威权主义:威权主义者信仰严格的社会等级制度,在所述社会等级制度中,他们完全服从其上级,并且预期其下级完全服从。由于严守规则,因此威权主义者人格是非常不安的,带有不确定性。

[0091] -自恋主义:自恋型人格具有强烈的自爱,以致其导致高水平的虚荣、自负和自私。自恋型个体通常难以与他人感同身受或对他人表示感激。

[0092] -自尊:积极地自我评估的倾向。自尊并不是暗示一个人相信他或她比他人更好,而仅暗示她/他是一位有价值的人。

[0093] -乐观主义:预期未来会有积极结果的倾向。乐观的人预期好事即将发生,并且通常会因此得到更多积极结果。

[0094] -述情障碍:无法识别和标记自己的情感。这些个体也难以识别他人的情感。

[0095] 要明确的是,这些人格特质是倾向。它们并不是决定性的(或“规范的”)。因此,在迈尔斯布里格斯术语中,ISTJ仔细检查他们购物车的可能性可能比ENFP高80%。这些倾向将在基础层级上影响IS将在各种社会情景中进行反应的方式。在一些实施方案中,可使用基本人格元素作为轴线在沿着多个轴线的连续统一体上设定IS的基本人格。

[0096] 例如且不限于,假设我们正创建一个名为“Chris?”的IS实体。我们可选择其性别和性偏好,因为这会影响人格,但还有更多会影响人格。使用迈尔斯布里格斯作为一种基本人格方法,我们可例如决定Chris是75%外向、25%内向;43%实感、57%直觉;27%思维、

73%情感;以及48%判断、52%知觉。类似参数可用于五因素方法。Chris可以:在经验开放性程度上为38%、在尽责性程度上为72%、在外向性程度上为81%以及在情绪性不稳定性程度上为22%。在另一个实施方案中,我们创建的人格可具有沿着其他连续统一体的程度,所述其他连续统一体像马基雅维里主义、成就感需要、认知需要、威权主义、自恋主义、自尊、乐观主义、述情障碍以及人格学家已知的其他连续统一体或由于AI行为分析而未揭示的其他连续统一体。

[0097] 所有这些都成为基本人格类型500,其是如图5所示的情景基线501的部分。此完整情景基线501是IS当下引入关系的内容。在基本人格类型之外还存在两个组分。首先,人们成长的环境会影响行为,即发展过滤器502。这包括可能已经影响了人自出生以来的发展的一切—与家庭的关系、此人在其中成长的文化、政治气候、甚至天气。人们也受他们与他们周围的那些人和事物的关系影响,即关系过滤器503。这包括与此人交谈的那些人以及此人与他们的历史。例如,关系的影响可取决于与他人的交互是否发生在办公室、度假村、某人的家中等。

[0098] 在一些实现方式中,人格类型的表示可具有16个基本人格组分并且具有与每个基本人格组分相关联的两个因素。因素1)是每个人格元素基于IS在一方面或另一方面的强烈程度在最大为100%的程度上的幅值,例如,他们是74%内向的、25%自恋的、17%判断的、49%马基雅维里主义的等。因素2)是16个人格组分中的每个人格组分在给定情景内的重要性的权重,例如,马基雅维里主义、思维或经验开放性对手头任务的重要程度。然而,本公开的方面不限于此类实现方式。

[0099] 此情景基线接着受到以下感觉输入504的影响:视觉505、听觉506、嗅觉507、味觉508和触觉509。这些输入接着由情感解读过滤器510进行解读,所述情感解读过滤器生成基线认知响应511。这些响应接着被馈送到行为函数算法512,所述行为函数算法512接着生成感觉输出513。

[0100] 发展过滤器

[0101] 根据本公开的方面,IS可被设计来模仿人类行为,这样做可创建IS的历史。当编写脚本时,他们通常具有描述是什么造就了角色的“宝典”。尽管脚本可能从未提及角色的出生地,但是了解到他们是在爱荷华州的农场还是在曼哈顿的联排别墅成长会极大地影响角色将采取的行动,并且因此影响演员将扮演该角色的方式。同样,发展过滤器是IS的角色宝典。例如,幸福的婚姻与不幸福的婚姻对一个人的行为的影响有很大的不同,并且在不幸的第一次婚姻之后拥有幸福的第二次婚姻对一个人的行为的影响仍是不同的。

[0102] 如图6所示,影响发展过滤器600的因素可分成几个关键储存桶。儿童期发展602包括以下事物,像家庭元素(规模、兄弟姐妹、父母等)、教育、财务状况、健康等。接下来是关于当前关系历史603的数据,我们还可追踪教育数据604、工作履历605和由本领域的专家确定为相关的任何数量的其他因素606。

[0103] 根据本公开的方面,可使用各种人工智能(深度学习、卷积神经网络、生成对抗网络等)来训练智能系统,并且可实质上即时地向所述智能系统给予其他机器学习,从而使得理解的深度将呈指数增长。可将任意数量的IS之间的交互与来自人类历史的实际交互进行比较并且进行微调。测试人员可为IS选择其他人格,并且运行其他人格以查看性能差异。应指出,不需要所有人类的完全准确的表示,几个文件化的人类交互历史可能就足够了。

[0104] 关系过滤器

[0105] 接下来,如图7所示,是IS 700和与其交互的一个人701或多个人702以及地点703、704、705的关系。已在此空间中存在许多心理学分析。根据本公开的方面,可根据心理学调查生成技术关系数据的语料库。心理学调查将回答以下问题,例如但不限于:一个人在涉及上司时与涉及下属时的感受如何(这将受到基本人格类型影响—他们是拘泥于礼节的人还是平等主义者)?其他家庭成员如何?一个人对遗传基因学(例如,一个他们从未一起长大过只是见过的亲戚)的感受如何?环境如何?他们在办公室环境中更舒适还是在酒吧或某人的家中更加放松?如果与他们交谈的人矮小或肥胖,他们会感到高人一等或自命不凡吗?他们有进行良好的目光接触吗?他们对身体提示的敏感程度如何?环境中的噪声如何?在一些实现方式中,调查可包括以下问题:要求调查接受者定义他们对每个问题的感受的权重和幅值。在其他实施方案中,从人类所回答的心理学调查的集合中统计得出权重和幅值。

[0106] 行为掩码

[0107] 在任何行为响应中,存在无限数量的响应。在相同情景中,不同的人将进行不同的反应。一些可能的影响领域在图8中示出。IS 801具有与人的关系802,所述关系包括以下事物,像业务联系、浪漫吸引力、对益智型兴趣或爱好的喜爱、他们在哪里成长、他们成长的家庭环境是什么、他们的健康史是什么、关系史是什么、他们的心理类型和倾向是什么?

[0108] 任何特定IS应被设计来掩蔽某些响应。尽管IS应被设计来掩蔽暴力,除非是在保护他人或也许自卫但仍不会造成伤害的情况下。这些关系过滤器或行为掩码是基于他们的生活史而概括的倾向,并且基于他们的背景和遗传基因学而具有普适性。一些背景可能是近期的,并且一些背景可能是更陈旧的,并且更深入、更根本的。

[0109] 行为函数

[0110] 如图9以高层级示出,在此我们采用前层;基本人格元素900、发展过滤器901、关系过滤器902和行为掩码903,并且将它们用作手头任务响应905的运算符。任务可以是例如但不限于;对对话中的问题做出响应、看向刚刚说过什么的某个人、决定是否购买商品、在第一选项不可用时选择另一餐厅或时间或日期、为购物者提供替代方案、或基本上人类今天可能会做出的任何回应。一个关键问题是“我们希望响应在何种程度上类人?”人类不是理性行动者,并且根据本公开的方面,IS可被配置来非理性地进行模仿。例如但不限于,假设某个人感到不安但对此无能为力。也许这个人只是错过了航班并且现在无法准时参加婚礼。理性行动者可能会说,对此无能为力了,已经探讨了所有选项,并且最佳的做法是发送一条道歉短信。然而,人类行动者可能会说:“天哪,真郁闷!让我们看看是否能够想到某一补救方式。首先让我们看看是否还有将使你及时到达那里的另一种方式。”在适当暂息和更多同情之后,他们可能会说:“你认为他们更喜欢你明早露面以在他们动身去度蜜月时告别还是仅仅寄个便条或发条短信?”关键是人类对过程与对结果同样关心。因此,IS可被设计来通过向适当的行为函数层904提供会提供过程意识的参数来模仿人类响应。

[0111] 作为另一个示例,假设一个人正在商店中为手表定价。人受制于称为“锚定”的认知偏见。作为一个示例,如果一件商品定价很高,则即使其他商品实际上仍然昂贵,此商品也会使这些商品看起来更加合理。如以下将论述,可开发在行为函数层904中具有适当掩码或过滤器的IS以对人类偏见做出响应。

[0112] 人类充满了认知偏见,这是使他们成为人的部分原因。人类坚信他们在做出选择

或确定价值之前会理性地分析所有因素,但事实是第一感受会在内心徘徊,从而影响未来的感受和决策。另一个认知偏见是“虚谈症”。人们坚信自己知道何时在自我欺骗,但是他们没有意识到人们从自己潜意识中形成的想法接收到的不断助推。行为心理学充满了以下研究,所述研究已经至少论证了以下命名的偏见:

[0113] 促发、虚谈症、后见之明偏见、德州神枪手谬误、拖延症、正常化偏见、自省、可得性启发、旁观者效应、达克效应、幻想性错觉、品牌忠诚度、诉诸权威、诉诸无知、稻草人谬误、人身攻击谬误、公正世界谬误、公共物品博弈、最后通牒博弈、主观验证、精神灌输、群体思维、情感启发、邓巴数字、售罄、自利性偏见、聚光灯效应、第三人效应、宣泄、从众、消弱突现、超常释放者、基本归因误差、代表性启发、社会惰化、习得性无助、锚定效应、自我实现的预期、契机、控制错觉、一致性偏见和预期。

[0114] 非限制的几个简短示例应足以论证这一般将如何起效。让我们以达克效应为例。达克效应是一种认识偏见,其中人们错误地夸大地评估了他们的认知能力。所述效应与虚幻的优越感的认知偏见有关,并且源于人们无法认识到自身的能力缺乏。对于设计为学习助手的IS而言,添加诸如达克效应的认知偏见使IS与用户更加相关,并且可能使学习更具娱乐性。例如,假设一个人正在学习以JavaScript进行编程,并且有一位专家程序员能够回答他的所有问题。然而,由于学习任何东西(包括编程)的乐趣都源自共享发现,因此这并不具娱乐性。带有达克效应偏见的IS将共享人的天真热情。这对于任何任务都将适用。通过举例而非通过限制,考虑学习打篮球。a) 与始终每投必中的某个人一起打篮球或b) 与强烈地意识到自身缺点的某个人一起打篮球将是不太有趣的。更有趣的是在开发能力方面要有一点人为鼓励和共享的天真信念,而不是一台能够冷淡地计算出你投中罚球或扣篮成功的实际几率的机器。

[0115] 后见之明偏见指代人们倾向于认为,相较于事件在其发生之前的实际可预测性,已经发生的事件具有更大可预测性。在此情况下,具有完美记忆力和感知力的IS将确切地了解它正确预测事件的频率,所述事件比如说天气会变坏。当然,人类基于感受到温度改变、也许是大气压力(“我的关节炎发作了”)等做出预测。预测明天将下雨而未下雨的人会忘记已经做出了预测,但如果的确下雨了,他会记得并且会说:“我就知道会这样!”

[0116] 图10示出了目前为止的整个堆栈。情景1000由情景基线(基本人格元素、发展过滤器和关系过滤器)加上行为掩码组成,并且这些项通过行为函数1001(诸如促发、虚谈症、正常化偏见等)进行过滤,并且制定响应1002。

[0117] IS可以做如下决策。认知偏见可被描述为一个函数,例如:行为 f (后见之明偏见) = (第1轴线上的已编程经验程度) (第2轴线到第 n 轴线上的已编程经验程度) (每条轴线上的事件固有可预测性)。例如,对于水手,已编程经验程度与水手对海边的天气轴线上的经验或水手对在沙漠中进行天气预报的经验有关。从IS系统角度来看,IS已经被“情景”促发一意味着其基本人格元素(Myers Briggs、基因组成、性别等),通过发展过滤器(文化和社会教养)修改,接着通过关系过滤器(与人们和环境的长期关系)情境化,通过行为掩码(当前与所涉及的人的关系、社会等级制度等)进一步修改。这样形成了函数(例如,行为 f (后见之明偏见))起作用的基本背景。

[0118] 在本公开的上下文内论述智能系统时,将认知偏见称为行为偏见更加有意义。将在稍后的章节中论述训练IS模仿这些行为偏见。

[0119] 在我们探讨将如何使用机器学习来训练IS之前,从高层级着眼于如图11所示的完整堆栈是有用的。对我们行为的最低层级输入是DNA 1100。DNA影响人格的方式将取决于专门针对IS生成的多个因素或与依照其塑造IS的人相关联的因素。DNA 1100的正上方是基于早期发展1101促成IS的方式。早期发展中发生的事情(虐待、极端贫困、全然的爱等)会非常深刻且通常永久性地对人进行塑造,并且也将会影响IS的人格。在同一时间框架内(并且持续程度更小)是基因调节1102,其通过基因和环境因素的组合来控制行为。上方是构成情景基线1103的元素—首先由基本人格元素1104构成,之后由发展过滤器1105和关系过滤器1106构成。从堆栈仍进一步向上爬升的是行为掩码1107,其分析行为偏见1108并且将那些偏见归咎于个体1109以及环境或行为和选择1110。根据所有这些数据,可使用行为函数1111来形成行为。此后,随着系统不断从其经验中学习,基于函数的交互的结果会被馈送回行为掩码中。

[0120] 人格面貌

[0121] 每个ISI都有成长或发展路径。此路径具有多个关键点,但一个点是它们变得不可替代的点—即当它们首次与人类交互时。例如且不限于,假设考虑个人(礼宾)客户服务代表,即称为Dale的IS。Dale了解跨个体客户的全部(任何制造商所制)装置的客户完整客户服务历史。Dale具有通过情景基线向上发展的人格。客户可从多个情景基线中进行选择,或可基于人格配置文件为他们选择一个情景基线。现在,向前进并且基于客户与IS的交互,他们的人格将得到发展。一年后,Dale将会了解客户认为什么是有趣的,客户是喜欢聊天还是只是开门见山,以及当然还有客户的所有购买和支持历史。这只是所述个体客户的人格面貌。如果另一人类客户以具有相同情景基线的代表开始,则“Dale”IS的所述客户的实例将不会长期保持不变。在IS与此客户交互时,所述关系将以与Dale与第一个客户的关系不同的方式发展。出于本公开的目的,通过情景基线(或我们的原始人格)向上发展的人格面貌在本文中称为基线人格面貌,并且“通过人际交互定制的”每个人格面貌都是ISI(智能系统实例)。如果两个客户结婚并且共享他们所有装置和个人客户服务合同,则这并不意味着他们必须共享支持人格面貌。一名配偶的支持人格面貌仍将是Dale,并且另一名配偶的支持人格面貌仍将是Alex(另一名配偶在两名配偶彼此相遇之前的支持人格面貌),但Dale和Alex都将有权访问两名配偶的所有总装置历史,只是根据谁进行呼叫,Dale或Alex将在电话线的另一端。本公开的方面包括以下实现方式,在所述实现方式中,配偶双方都在电话会议中,其中Dale和Alex同时都在电话中并且人格将自然混合。因此,IS为你和你的妻子形成“增强的”客户支持体验。

[0122] 根据如图12所示的本公开的用于数据管理的方面,存在基线人格面貌的存储库1200或用于动态形成基线人格面貌的元素,并且在每种新情景中,从一个基线人格面貌形成IS实例1201。此实例可被存储并在每次交互之后进行更新,或它可基于先前交互的参数每当需要时动态地重新形成。在一些实施方案中,IS实例被高速缓存有限的时间段以消除延迟,但参数被存储,以使得即使它们已离线较长时段,它们也可精确地在其停止的地方重组。

[0123] 训练智能系统

[0124] 机器学习(深度神经网络、机器学习、CNN、RNN、GAN等)可被实现来捕获这些行为偏见并对其进行分类,然后在“扮演人类”时模仿这些行为偏见。着眼于图13,以上已经论述基

本人格的一些其他层1300:发展过滤器、关系过滤器和行为掩码。映射人类认知行为的下一个层是用于映射行为偏见。为了训练IS扮演人类,存在可采用的多个步骤。行为集合1301开始于由心理学家基于人们对认知偏见1302已有的了解构建的专家系统。这由人类世界和人/IS虚拟世界中的可观察到的行为数据1303增加、增强并主要由其替换。可观察到的行为数据可通过观察人类世界中的对话而生成。基于人的认知偏见和心理学配置文件的模型,存在对人在某些对话环境中将如何反应的预期,并且当反应与模型不同时,更新所述模型。在此情景中,对话环境可通过以已知心理被动地观察人类之间的对话而生成,或通过人类与IS之间的对话而主动地生成。IS可向具有已知心理学配置文件的人类提供话题或与此人讨论话题,并且基于预测的响应来判断人类响应。可基于实际的人类响应来更新预测的模型。接下来,将行为分析映射到认知偏见1304。所得的行为偏见用于对以下进行归咎:IS在其为某种类型的个体(基于以上所有层)时响应的方式1305;以及同样那些偏见适用于不同行为和选择的方式1306。将个体行为预期1305和环境选择1306的组合应用为函数以形成IS的行为偏见1307。可观察函数1308的行为和运作,并且将该学习馈送回可观察到的行为数据1303中。一旦IS的实例工作,他们就可利用GAN(生成对抗网络)开始相互训练以继续演变。

[0125] 广义神经网络训练

[0126] 根据本公开的方面,IS系统可包括若干不同类型的神经网络中的一者或多者,并且可具有许多不同层。通过举例而非通过限制,分类神经网络可由一个或多个卷积神经网络(CNN)、递归神经网络(RNN)和/或动态神经网络(DNN)组成。

[0127] 图29A描绘具有一层节点2920的RNN的基本形式,每个节点层的特征在于激活函数S、一个输入权重U、回归隐藏节点转移权重W和输出转移权重V。激活函数S可以是本领域已知的任何非线性函数,并且不限于(双曲正切(tanh)函数。例如,激活函数S可以是Sigmoid或ReLU函数。与其他类型的神经网络不同,RNN对于整个层具有激活函数和权重的一个集合。如图29B所示,RNN可被视为移动通过时间T和T+1的具有相同激活函数的一系列节点2920。因此,RNN通过馈送从先前时间T到当前时间T+1的结果来维持历史信息。

[0128] 在一些实施方案中,可使用卷积RNN。可使用的另一种类型的RNN是长短期记忆(LSTM)神经网络,它在RNN节点中添加具有输入门激活函数、输出门激活函数和遗忘门激活函数的存储块,从而得到门控存储器,所述门控存储器允许网络保留一些信息达更长时间段,如1997年《神经计算》第9卷第8期第1735至1780页的Hochreiter和Schmidhuber的“Long Short-term memory”所描述,所述文献以引用方式并入本文。

[0129] 图29C描绘根据本公开的方面的诸如CRNN的卷积神经网络的示例性布局。在此描述中,针对输入2932生成卷积神经网络,所述输入具有以下大小:高度为4个单位以及宽度为4个单位,从而给出16个单位的总面积。所描绘的卷积神经网络具有高度为2个单位以及宽度为2个单位的过滤器2933大小,其中跳跃值为1并且通道2936的大小为9。在图2C中,为简洁起见,仅描绘通道的第一列与其过滤器窗口之间的连接2934。然而,本公开的方面不限于此类实现方式。根据本公开的方面,实现分类2929的卷积神经网络可具有任何数量的另外的神经网络节点层2931,并且可包括此类层类型作为另外的卷积层、完全连接的层、池化层、最大池化层、任何大小的局部对比度归一化层等。

[0130] 如图29D所见,训练神经网络(NN)从NN的权重的初始化开始2941。通常,初始权重

应随机分布。例如,具有 \tanh 激活函数的 NN 应具有在 $-\frac{1}{\sqrt{n}}$ 与 $\frac{1}{\sqrt{n}}$ 之间分布的随机值,其中 n 是节点的输入的数量。

[0131] 在初始化之后,定义激活函数和优化程序。接着向 NN 提供特征向量或输入数据集 2942。NN 可根据具有已知标签的输入生成不同特征向量中的每一者。类似地,可向 NN 提供对应于具有已知标签或分类的输入的特征向量。接着,NN 预测特征或输入的标签或分类 2943。将预测的标签或分类与已知标签或分类(也称为真实值(ground truth))进行比较,并且损失函数测量所有训练样本的预测值与真实值之间的总误差 2944。通过举例而非通过限制,损失函数可以是交叉熵损失函数、二次成本、三元组对比函数、指数成本等。可根据目的使用多个不同损失函数。通过举例而非通过限制,对于训练分类器,可使用交叉熵损失函数,而对于学习预训练的嵌入,可采用三元组对比函数。接着,使用损失函数的结果以及使用训练神经网络的已知方法(诸如使用自适应梯度下降的反向传播等)对 NN 进行优化和训练 2945。在每个训练时期中,优化器尝试选择使训练损失函数(即总误差)最小化的模型参数(即,权重)。数据被分成训练样本、验证样本和测试样本。

[0132] 在训练期间,优化器使训练样本的损失函数最小化。在每个训练时期之后,通过计算验证损失和准确性来评估验证样本上的模型。如果没有重大改变,则可停止训练,并且可使用所得的训练模型来预测测试数据的标签。

[0133] 因此,可根据具有已知标签或分类的输入训练神经网络以对那些输入进行识别和分类。类似地,可使用所描述的方法来训练 NN,以根据具有已知标签或分类的输入生成特征向量。

[0134] 生成对抗 NN 训练

[0135] 训练生成对抗 NN (GAN) 布局需要两个 NN,如图 30 所示。两个 NN 彼此相对设置,其中第一 NN 3002 根据源响应 3001 和目标响应 3005 生成合成源响应 3005,并且第二 NN 将响应 3006 分类为是或不是目标响应 3004。基于第二 NN 3006 做出的分类来训练 3008 第一 NN 3002。基于分类是否正确地识别目标响应 3004 来训练 3009 第二 NN 3006。第一 NN 3002 (在下文称为生成 NN 或 G_{NN}) 采用输入响应 (z) 并且将其映射到表示 $G(z; \theta_g)$ 。

[0136] 第二 NN 3006 在下文称为判别 NN 或 D_{NN} 。 D_{NN} 采用未标记的映射合成源响应 3006 和未标记的响应 (x) 集 3004,并且尝试将这些响应分类为属于目标响应集。 D_{NN} 的输出是单个标量,其表示响应是来自目标响应集 3004 的概率。 D_{NN} 具有数据空间 $D(x; \theta_d)$,其中 θ_d 表示 NN 参数。

[0137] 在生成对抗 NN 的训练期间使用的 NN 对可以是多层感知器,所述多层感知器类似于上述卷积网络,但每一个层都是完全连接的。生成对抗 NN 不局限于多层感知器,并且可被组织为 CNN、RNN 或 DNN。另外,对抗生成 NN 可具有任何数量的池化或 softmax 层。

[0138] 在训练期间, G_{NN} 3002 的目标是使 D_{NN} 的逆结果最小化。换句话说,训练 G_{NN} 以使 $\log(1 - D(G(z)))$ 最小化。在训练初期可能会出现以下问题: D_{NN} 拒绝具有高置信水平的映射输入响应,因为它们与目标响应集有很大不同。因此,方程 $\log(1 - D(G(z)))$ 迅速饱和,并且学习变慢。为了首先克服这个问题,可通过使 $\log D(G(z))$ 最大化来训练 G ,所述 G 在学习早期提供强烈得多的梯度,并且具有相同动态固定点。另外,如 Zhu 等人的可在 <https://arxiv.org/pdf/1703.10593.pdf> (30Aug 2018) 处获得的“Unpaired Image to Image Translation

using Cycle-Consistent Adversarial Networks”ArXiv,ArXiv:1703.10593v5[cs.CV]所论述,可将GAN修改为包括循环一致性损失函数以进一步改善映射结果,所述文献以引用方式并入本文。

[0139] 训练 D_{NN} 3006的目标是使向训练数据集分配正确标签的概率最大化。训练数据集包括映射源响应和目标响应两者。 D_{NN} 提供标量值,所述标量值表示训练数据集中的每个响应属于目标响应集的概率。因此,在训练期间,目标是使 $\log G(x)$ 最大化。

[0140] 第一NN和第二NN一起形成双层极小极大博弈,其中第一NN3002试图生成响应以欺骗第二NN 3006。所述博弈的方程是: $\min_G \max_D V(D,G) = E_{x \sim p_{数据}}(x) [\log D(x)] + E_{z \sim p_z}(z) [\log 1 - \log D(G(z))]$

[0141] 通过优化 D_{NN} ,然后优化 G_{NN} 以逐步方式训练 G_{NN} 和 D_{NN} 。重复此过程众多次,直到在鉴别器中看不到进一步改善为止。当训练响应是映射输入响应的概率 p_z 等于训练响应是源响应的概率 $p_{数据}$ 时,就会发生这种情况。换句话说,当 $p_z = p_{数据}$ 、替代地 $D(x) = 1/2$ 时。总体上类似于以上针对神经网络论述的内容,可使用小批量随机梯度下降法或用于训练兼容神经网络的任何其他已知方法来训练 G_{NN} 和 D_{NN} 。有关对抗生成神经网络的训练和组织的更多信息,请参见Goodfellow等人的在<https://arxiv.org/abs/1406.2661>处可获得的“Generative Adversarial Nets”arXiv:1406.2661。

[0142] 专家系统

[0143] 专家系统通常使用正向链接或反向链接。根据本公开的方面,专家系统的一些实施方案可使用正向链接。另外,本公开的实施方案可使用前景理论并且生成合成数据集以辅助专家系统的发展和训练。如图14中可见,最初存在认知偏见集1400。尽管所述领域在不断扩展,但目前存在47种常见认知偏见:促发、虚谈症、确认偏见、后见之明偏见、德州神枪手谬误、拖延症、正常化偏见、自省、可得性启发、旁观者效应、达克效应、幻想性错觉、品牌忠诚度、诉诸权威、诉诸无知、稻草人谬误、人身攻击谬误、公正世界谬误、公共物品博弈、最后通牒博弈、主观验证、精神灌输、群体思维、超常释放者、情感启发、邓巴数字、售罄、自利性偏见、聚光灯效应、第三人效应、宣泄、误导信息效应、从众、消弱突现、社会惰化、透明度错觉、习得性无助、具身认知、锚定效应、注意力、自我设限、自我实现的预期、契机、一致性偏见、代表性启发、预期、控制错觉和基本归因误差。应指出,上述列表不受限制地提供,因为已知偏见的领域目前正在扩展,并且可利用本领域普通技术人员已知的足够的编程来添加任何数量的偏见。这些偏见已由专家确认1401。接着(仍使用专家)将基线值归因于每个行为1402。接着将这些基线值存储在基线偏见的专家系统量化的数据存储区1403中。最终,这是否完全准确并不重要,因为系统将随时间推移从实际行为中学到更多,但是必须有一个基线来开始所述过程并且给我们一个通过其查看行为的镜头。

[0144] 专家系统的发展开始于:确定认知偏见并将其映射到情景环境、以及将基线值归因于每个行为1402。例如且不限于:为了获得一些幻想性错觉(错误地感知无关事物之间的关系和含义的倾向)的基线数,心理学家着眼于巧合(例如偶遇你认识的某个人或发现班上的那个人与你的生日相同),并且列出最常见的100种巧合。接着,心理学家通过统计和实验数据确定每个事件的可能性相对于可能的惊讶程度的权重。此统计和实验数据可通过对受试人的调查或观察来生成。这些映射是基线映射—根本不基于任何个体人的人格,而只是典型的预期反应。关于专家系统的好处在于,它们可自然加权,使得结果不局限于二元结果。

例如且不限于,着眼于遇见从家乡出来度假的某个人。由于社会联系、共同媒体、共同社会行为、气候和假期安排,偶遇与自己共有共同好友的某个人可能不是那么罕见,但其始终是令人惊讶的。例如,在拉斯维加斯的大型度假村,概率可能暗示在春假期间,这种情况每30次社会交互将发生一次,但当受到询问时,人们已经暗示他们将预期这种情况每100次发生一次。因此,对于此特定幻想性错觉情况,其将被给予“过度可能性值”3.33 (100/30)。将收集一些数量(也许是几十个)幻想性错觉示例,并且根据所述示例生成最常见情景的幻想性错觉基线集。这些与我们的所有其他行为偏见基线数据存储在—起。接着,当专家系统遇到幻想性错觉将应用于人类的社会情景时,专家系统将适当的基线应用于所述情景。我们可使用针对每个行为偏见定义专家系统基线的类似方法。

[0145] 形成基线人格

[0146] 为了计算对基线人格的影响,必须首先存在一种表示基线人格状态的方式。为此,使用一系列矩阵,其中每个矩阵是社会学或心理学领域内的多个维度,并且每个层是作用于其下一层的掩码或函数。图15描绘构成人格基线1500的组分的框图。存在:基础层1501,其由DNA、RNA、性别、身体素质、Myers Briggs、5因素、人格特质构成;文化层1502,其由养育、国家、州、城市、街区、宗教、文化、家庭结构等构成;训练层1503,其由早期学习环境、儿童保育、学习焦点、教育、工作经验等构成;一般环境层1504,其由国家、城镇、自然环境等构成;以及特定环境层1505,其由社会环境、天气、当日时间和其他相关因素构成。

[0147] 关于每个矩阵内的数据的鲁棒性,描述未必是完全准确的。如果某个人被描述为在自恋主义程度上为13%,则这是否是一个准确数字并不重要,因为数字不是重点。目的是得到人格结果,并且因此系统将随时间推移学习以基于实际行为结果来量化人格因素,并且将随时间推移来调整构成矩阵、掩码和函数的各种参数的权重。训练集首先是如专家系统描述的行为分析。专家系统是线性的,其中正向链中的每个链接源于其前因变量。然而,在心理学系统中,存在许多因素,并且它们未必是确定性的。实际上,数百种其他推论因素可与每个预测行为相关联。例如且不限于,自恋主义的组分可以是:1)有夸大的自负感;2)有权利感并且需要持续的过度赞美;3)预期被认为是优秀的,而甚至没有证明优秀的成就;4)夸大成就和天赋;5)全神贯注于关于成功、权力、才华、美貌或完美配偶的幻想;6)坚信自己是优秀的,并且只能与同样特殊的人交往;7)独占对话并且贬低或看不起他们视为下等的人;8)预期特殊恩惠并且盲目地依从他们的预期;9)利用他人来获得他们想要的东西;10)无法或不愿认识到他人的需要和感受;11)嫉妒他人并且坚信他人也嫉妒自己;12)表现得傲慢或高傲,给人留下自负、自夸和自命不凡的印象;13)坚持拥有最好的一切。如果一个人与所有这些组分都有很强的关联,那么他们肯定是相当自恋的。然而,如果他们对于组分中的一半是适度的,则他们可能仍是自恋的,但自恋程度较小。可对具有已知心理学配置文件的人(在此情况下为13%的自恋的人)执行对话分析。使用如最初由心理学家定义的提示集,将根据这些对话确定典型的响应和提示。因此,构建构造好的原则集,所述原则集权衡关于自恋程度的决策。构建所有其他心理学组分的类似列表。可观察到接近13%的自恋者的所有已知人格,并且确定哪些其他因素构成他们的人格,并且当其他人格特质被发现时,所述其他人格特质被映射回自恋的量度。

[0148] 接下来,因素组可用作神经网络的输入向量。使用机器算法预测行为集的标签(其中所述标签是基于人格量度的程度)来训练的神经网络可用于基于人格量度的程度来标记

因素组。接着使用形成这些因素的语言(例如,我们原始训练集中的cat)来确定哪种语言会形成关于心理学定义的倾向。由此,可生成对应于不同心理学配置文件的语言。

[0149] 基础层

[0150] 基础层的第一组分是DNA。人的全面拟像包括许多(如果不是全部的话)影响人的人格的因素,所述因素包括基因组成。根据本公开的方面,基础层的DNA可被表示为影响人格的重要已知基因序列或表示为影响人格的基因注定的条件。基础层中的DNA信息可能例如且不限于可以是影响人格的因素,像身体性别和性别认同、体型、协调性、视觉和听觉敏锐度以及其他物理原语(像心脏病或糖尿病的倾向)。也存在心理物理原语,像阅读障碍和左撇子。DNA因素可以是基础层矩阵的第一维度。这些基因基于RNA对它们的转录和调控可能表达或可能不表达自己,并且因此RNA形成我们的基础层矩阵的下一个维度。RNA随时间推移以不同方式表达自己,并且因此具有动态效应—主要是在人生的早期阶段。另外,在早期阶段期间是社会学影响,一些社会学影响在发展的非常早期,像母乳喂养或睡眠训练。与DNA和RNA有关的矩阵的维度可由遗传学家定义,并且可随着有关DNA对人格影响的信息增加而改变。非常早期发展的矩阵的维度可由儿童早期心理学家定义。应指出,矩阵中的每个条目都必须进行加权。同样,我们可基于本领域专家的意见从基本权重开始,但最终的权重及其对人格的影响将基于观察和经验随时间推移而更新。基础层矩阵的下一个维度是人格连续统一体,如图16所示。如上所论述,在根据本公开的方面的一些实现方式中,沿着外向->内向1601、实感->直觉1602、思维->情感1603、判断->知觉1604的轴线的加权矩阵1600中的迈尔斯布里格斯类型指标(MBTI)百分比为我们的基础层矩阵的维度提供数字集。另一个维度可以是五大人格特质:经验开放性、尽责性、外向性、亲和性和情绪不稳定性。存在由其他人格连续统一体表示的又一个维度,诸如马基雅维里主义、成就感需要、认知需要、威权主义、自恋主义、自尊、乐观主义和述情障碍。所有这些因素加在一起可用于形成基础层矩阵。这是最基础层级处的人格表示。下一个层是文化层。

[0151] 文化层

[0152] 文化层捕获IS被设想为已成长的方式。此层包含有关IS的背景信息,如图17所示,诸如且不限于;IS是在哪个国家、州、城市和街区长大的1701?IS是在城市地区还是在农村地区长大的1702?IS与什么宗教有联系1703?IS成长的政治气候是什么1706?IS成长所处的文化是什么1704,并且什么家庭结构1705是其教养的部分。

[0153] 跨文化存在很多多样性。“亚洲文化更为集体主义,并且这些文化中的人们趋于不太外向。中美洲和南美洲文化中的人们在经验开放性上趋于得分更高,而欧洲人在情绪不稳定性上得分更高。”这些差异全部被文化层捕获,并且促成了IS中的不同人格。

[0154] 作为文化对人格影响的另一个示例,生活在集体主义文化中的人们(通常是亚洲、非洲和南美洲)将社会和谐、有礼和群体需要视作高于个体需要,而个人主义文化中的人们(通常是北美洲、澳大利亚、欧洲)显示出更多个人主义人格特质。

[0155] 美国内部似乎也存在地区性人格差异。研究人员分析了来自美国超过150万个体的响应,并且发现三个截然不同的区域人格群集:“友好和传统……;更加放松、情感稳定、平静和富有创造力……以及充满压力、急躁和沮丧。”

[0156] 这些文化数据适合矩阵层块。根据本公开的一些实施方案,不需要表示所有城市和村庄以及所有家庭。只需要足够数量的示例来形成我们的人格。例如且不限于,为了形成

客户服务IS实例,只要几个代表性背景就足够了,例如人口在200,000与1,000,000之间的100个美国城市—不需要来自所述大小的所有美国城市的代表性样本。着眼于文化教养层块,描述可例如且不限于开始于场所,从不太精细到更精细:大陆→国家→县→镇→街区。心理学和社会学的学生可将行为倾向映射到这些层中的每一层上。因此,例如,基于收集的MBTI数据,尼加拉瓜人的外向度比平均水平高15%,而智利人的内向度比全球平均水平高11%。下一个层是城市↔农村。在城市-郊区-农村之间以及它们之间的空间之间存在已知的不同态度差异。这些都经过编码—同样作为专家系统开始,结合对真实对话(文本、电话、电影等—参见下文)的分析,并且随时间推移,它们基于来自用户的反馈并且基于与具有相似想法的其他人的密切关系进行微调。

[0157] 使用类似方法,可添加其他各种文化维度:宗教、文化教养、家庭结构和政治环境。这些同样由社会学家和心理学家来排名,但将通过电影和电视节目、聊天会话、文本消息、语音消息和电子邮件的分析来了解。

[0158] 训练层

[0159] 在此背景下,训练称为IS从很小的时候就开始的教育或学习。此层将回答有关IS的问题,诸如且不限于:早期学习环境(从母乳喂养和目光接触以及阅读给孩子听到儿童保育)的影响是什么。随着他们年龄的增长,学习环境是什么?例如,是男女同校吗?大班还是小班?是妨碍的还是专注的?高等教育和工作经验如何?我们的IS实例是如何训练的?她/他上过大学吗?他们的专业是什么?他们加入过兄弟会/姐妹会吗?他们的成绩是多少?研究生学位或证书怎么样?先前工作履历怎么样?所有这些元素都会支撑我们的IS实例的人格。本质上,为我们的IS实例形成虚拟简历。训练层不必形成IS要记住的实际事件(例如“记得1995年那个盛大的AEPi万圣节啤酒狂欢,Jeffery喝的太多了”)。训练层足以形成确保独特类人人格的丰富宝典。在一些实现方式中,角色宝典可由用户或软件程序员利用允许创建者为IS编写独特历史的工具来形成。在其他实施方案中,IS基于情景利用它们自己的宝典形成他们自己的实例。在又其他实施方案中,可构建混合模型,其中可概述显著特征,并且IS将提供选择以供从中进行选择—例如且不限于可选择IS实例的宝典中的分支,同时IS的底层原型不可变。

[0160] 一般环境层

[0161] 一般环境层描述IS当前状态的背景。换句话说,它描述导致IS与用户联系的事件。此层可回答诸如且不限于以下问题:此IS在哪里工作?影响人生的早期学习阶段的相同因素中的一些因素会影响工作或娱乐环境。IS实例是在呼叫中心、在酒吧中、在办公隔间(cube)中、正打销售电话还是在律师事务所中作为诉讼律师正在工作?在一些实施方案中,一般环境层可开始于以下基本职业分类:农业、食品和自然资源;建筑和工程、艺术、音频/视频技术和通信;商务管理与行政;教育和培训;金融;政府和公共行政;健康科学;酒店和旅游;人类服务;信息技术;法律、公共安全、惩教和安保;制造;营销、销售和服务;科学、技术、工程和数学;以及运输、分配和物流。接着,覆盖在其上的是在教养背景下使用的相同文化和背景变量:大陆→国家→县→镇→街区;城市→郊区→农村。接着可添加另外的分支,诸如大公司↔小公司、大建筑物↔小建筑物;劳动力的(在多个轴线上的)多样性程度;公司/行政部门的政治/宗教信仰是什么;等等。分类的域和子域可由社会科学家定义,但类别、其权重和重要性都取决于与用户联系的结果以及那些输入与那些输出的值之间的相关

性。

[0162] 特定环境层

[0163] 下一个层是特定环境层。这里，捕获用于形成IS实例行为的最新元素。此层可回答诸如但不限于以下的问题：天气如何？交通如何？我早上过得好吗？例如，基于我们ISI的构建家庭，他们可能处于大体上幸福的婚姻中，这段婚姻拥有必须去上学的两个孩子，并且处于意外事件（孩子生病、家庭作业丢失等）的正态分布下，我们的ISI的情绪以及因此其行为受到这些预备因素影响。

[0164] 行为分析

[0165] 训练的下一个任务是让IS学习人类行为。存在许多交互式语音响应（IVR）系统，其使用自动语音识别（ASR）技术和自然语言处理（NLP）来分析现今市场中的人类行为。这里的方法是其用作对话基线。根据本公开的方面，IS被配置来带来不仅是即时背景（此人在购物或在海滩等），而且还带来个人和社会背景—了解人类社会行为。

[0166] 图18描绘根据本公开的方面的行为集合1800和权重的概述。最初，专家系统可利用从我们的专家系统收集的认知偏见1801来促发，并且可添加可观察到的行为数据的语料库1802。本公开的实施方案可使用聊天机器人数据的语料库1803并且分析更多正常/社会对话环境。这些可非限制性地包括；电子邮件和文本1804、社交媒体1805、语音邮件1806、电影、电视节目和流式传输视频源1807。视频媒体在社交方面非常丰富，并且存在可访问的大量丰富的数据。存在许多电影和电视节目不典型或“标准”的风险，并且因此我们将采用人类AI分类器1808（例如但不限于土耳其机器人或人类AI分类器的各种其他网络）来在常态程度上对对话进行评分。稍后，当IS与人聊天时，它可通过人类响应进行学习。另外，它可与其他IS实例聊天（使用GAN），并且我们可雇用同一人来分析对话。根据通信数据的特定语料库，我们将需要多种不同方法来进行分析。接着，我们将把行为分析映射到行为偏见1809。

[0167] 根据本公开的方面，电影和电视可以是用于训练IS的信息的一个源。这些未必表示典型的长期弧。许多体裁始终有一个幸福的结局，其他体裁则是尖锐批评的，情景喜剧通常依赖于人们撒谎，然后喜剧随之而来。然而，在许多这些相同标题中，一分钟接一分钟的行为是非常人性化的。妙语或令人震惊的事件通常令人惊讶并且并不典型，但动作之间的所有现象都是正常的。在一些实施方案中，向土耳其机器人/分类器馈送数以万计的视频资产以针对常态进行分级。此方法适用于电影和电视，但还有其他视频资产—最值得注意的是流式传输社交媒体视频（例如YouTube、Facebook视频等）。在一些实施方案中，神经网络1808用于执行社会分析并且过滤“非正常”行为。根据定义，此语料库的价值几乎与其他社会分析的价值不同，并且因此必须使它们（以及实际上每个分组）保持分离（标记为自己的语料库）。

[0168] 数据的下一个语料库是聊天机器人数据。当人们问问题、寻求具体答案时，这将尤其相关。来自各种聊天机器人语料库的响应将集中在准确性上。根据本公开的方面，IS可被训练成未必提供正确答案，而是提供最人性化的答案。从聊天机器人数据中得出的一个有价值的结果是当机器人理解错误时。因为正在询问机器人的是真正的人类，所以他们将并不总是满意“正确的”答案（暂时忽略机器人误解或曲解问题的次数）。因此，在一些实施方案中，如果客户由于软件未被设计来解决他们尝试解决的问题而遇到困难，则响应将通常对人类来说是不令人满意的，而不是说：“对不起，但是我们的软件确实做不到那样。”更灵

敏的机器人可能会说：“让我看看我是否恰当地理解了您尝试解决的问题。”接着在恰当地确认机器人“理解了”问题之后，他们可能会将所述问题交至恰当软件以解决他们的问题（有点像34街的奇迹里的梅西百货的圣诞老人，其让家人参与竞争，从而大大增加了梅西百货品牌的价值）。也许机器人甚至是“他们的机器人”，也就是说，他们陪着客户以帮助他们恰当地安装他们的其他软件（实际上，相同机器人人格面貌将随后参与竞争，并且客户数据的相关部分将在允许之下进行共享）。

[0169] 用于行为分析的语料库列表上的下一个可以是社交媒体分析。此分组内存在大量通信数据，并且通常对所述通信数据进行组织。人们通常试图对他们的相识之人表现出其最好的一面，并且对他们的亲密朋友表现出其更加真诚的一面。匿名或伪匿名帖子可能特别不典型，并且可能根本不应包括在内。私人帖子由于通常与其相关联的自然坦率而将具有特定价值（像“另一个完美的一天”或“生活一如既往地糟糕”的无谓的帖子会被过滤掉）。同样，将私人帖子储存到多个分开的语料库中，并且通过分析对行为的实际响应来定义人性的权重。

[0170] 语料库的最后的储存桶是文本、语音邮件和电子邮件。就语气而言，文本是最自然的。在一些实施方案中，表情符号和首字母缩写词被扩展为由表情符号或首字母缩写词表达的情感或背景的真实语言描述。了解各种聊天的成员之间和当中的关系将为每个组提供背景或元数据。一个人对其母亲讲话的方式将不同于对其朋友讲话的方式，并且对其朋友讲话的方式将不同于对一组4个朋友讲话的方式。此背景的应用对于我们的AI训练将非常有价值。在某些方面，语音邮件是文本数据的子集。它通常针对单个人或住户。用户或语音邮件通常是与文本消息的人口统计资料不同的人口统计资料，并且可表示不同或更旧集的人格。如今的年轻人很少使用语音邮件。电子邮件可能需要元数据来描述情景并且分成多个语料库。在一些实施方案中，线程可被解析为不同语料库（例如且不限于，商务电子邮件内的个人通信）。

[0171] 下一个任务是使上述行为观察与行为倾向（行为偏见）列表相关1809。此解析回答了以下问题：哪些行为与哪些倾向相关联。在这里，专家分两个阶段使用。首先，专家将行为分成组。例如且不限于，脾气暴躁的、同理的、焦虑的或放松的人们。专家形成情感分组的分类，并且然后人分类器标记所有行为。接着将这些行为及其标签用于训练机器学习系统（即神经网络）。

[0172] 将行为偏见归咎于个体

[0173] 如今，每个人对每种行为偏见所表现的程度不同。一个人可能是非常“善于分析的”和“现实”的，而另一个人可能更情绪化的并且易于夸张。即使当易于夸张时，这通常局限于一些域。夸大他们在其高中篮球队中的角色的同一个人当想起他们去希腊旅行的细节时可能是非常准确的。请记住情景基线本质上是从堆栈向上发展的IS的人格（DNA⇒早期发展+基因调控⇒基本人格元素⇒发展过滤器⇒关系过滤器）。

[0174] 如图19所示，映射针对行为偏见1901中的每一者的情景基线1900以形成每个IS实例的情景偏见集1902。通过举例而非通过限制，开始于一个IS实例，并且然后外推到更一般的实例。例如，可利用参数生成以下IS实例：在中西部小镇成长的35岁异性恋女性。她成长为新教徒，但并未定期去教堂做礼拜一然而，她有强烈的公平价值观。她有两个在世的兄弟姐妹，一个哥哥和一个妹妹。她的哥哥是同性恋，并且这影响她对关系的开放态度（不武

断)。然而,她关于堕胎和生存权问题是武断的,并且坚信有足够多的人想要孩子并且因此不应废弃任何生命。她是EFTJ(64、81、59、50, Myers Briggs),并且她的5个因素是开放性(66%)、尽责性(34%)、外向性(72%)、亲和性(48%)和情绪不稳定性(32%),并且她的其他人格特质为情绪不稳定性(53%)、马基雅维里主义(14%)、成就感需要(21%)、认知需要(73%)、威权主义(26%)、自恋主义(14%)、自尊(58%)、乐观主义(87%)和述情障碍(13%)。她喜欢填字游戏和医疗剧。她的角色宝典潜在地存在更多组分,但这对于IS作为虚拟人是谁给出了一定想法。现在如图20所示,我们可将所有心理学参数映射到多个行为偏见2000,并且向每一者提供权重2001。这些是基线偏见,并且尚未被映射到任何具体情景。然而,它们各自具有基线值,也就是说,“你的行为偏见有多明显?”如上所论述,我们将利用许多心理学家将Myers Briggs和五因素倾向映射到我们的行为偏见。这可在任何程度上完成,但是通过举例而非通过限制,可使用百分比,其中更高的数字更易于发生偏见以及低数字较不易于发生偏见。

[0175] 随着IS与其他用户和ISI交互,IS变得更加成熟,偏见的数量、名称和权重将发生改变。即使从初始偏见集开始,ISI具有其将如何进入新的情景的基本人格因素集。现在我们可以开始将带有偏见的IS应用于具体情景。请记住行为偏见是人格的仅一个组分,并且因此在我们准备解决具体情景时,需要完整的人格宝典。例如,上述异性恋女性对堕胎和收养的敏感度将强烈地影响她在某些情景中的举止,并且她可能不是计划生育早期妊娠支持小组的正确人格,但她对于基于教堂的早期妊娠支持小组可能是正确的形象。

[0176] 将行为偏见归咎于情景环境

[0177] 图21示出堆栈中的下一个层:将IS实例映射到它所处的情景。映射情景基线2100和行为偏见2101以形成每个IS实例的情景偏见2102,并且接下来我们将把偏见归咎于情景环境2103。

[0178] 此层回答了以下问题,诸如且不限于IS是在化妆品柜台销售化妆品、提供客户支持、充当保健人员还是度假村的顾问?在这里,一些情景行为数据是有用的。这可从具有基于心理学数据的轴线的专家系统开始。如从图22可看出,存在在最高层级上基于地域2200(西伯利亚与戛纳有很大不同)的行为参数。每个人还在某种程度上受到他们所处的当日时间2201影响。许多其他参数是两个不同方面之间的连续统一体。例如且不限于,城市社会2202与农村社会2203之间存在较深的社会差异。为此,我们可使用人口密度。还存在有权2204与朴实2205之间的连续统一体。人们在比佛利山庄以一种方式行事,并且在康普顿以另一种方式行事。环境有多拥挤?IS是在繁忙的购物中心还是单独在自己的家中工作(这是经由建立人格行为进行的一它实际上未必具有任何物理实例化,在一些实施方案中,IS将“生活”在云端)?其他因素将包括但不限于在大型群体2206中或独自2207工作或娱乐、他们从事何种工作2208以及从事何种娱乐活动2209、他们的工作环境如何2210、以及他们在空闲时去哪里闲逛2211。现在当我们着眼于各种环境时,我们可基于类型略微进行深入研究。是保险公司吗?是图形设计公司吗?是软件程序员办公隔间吗?是图书馆还是在沙滩上还是在游轮上?我们将使用人类创造的专家系统来形成完整分类。所述系统是否完全准确并不重要,因为AI将随时间推移而迭代,并且基于经验开发自己的分类。

[0179] 一旦根据心理学和社会学数据设定我们专家系统的基本参数,则分类器可将预期的行为映射到社会环境。在此情况下,收集大量感知的社会习俗比确定社会规范的准确表

示更为重要。这可通过对大量人口进行调查或以其他方式进行投票来完成。在足够数量的人(例如,约一千人)已经对具体行为和环境进行了调查之后,系统将具有相当准确的社会视角。

[0180] 通过举例而非通过限制,假设IS是银行柜员。现在我们尚未与个体进行具体交互,但是我们确实了解有关环境的多个事项。假设我们的柜员在一家人们趋于友善和社交的农村银行。银行通常并不拥挤,并且因此人们通常不必等待,并且交易通常从有关天气或有关某一最近事件或集市的对话开始。让我们假设柜员有相当长的(虚拟)通勤时间,并且将了解当天的交通情况。所有这些因素都会支撑此环境中的IS实例的基线行为。

[0181] 对数据进行排序

[0182] 此堆栈中的下一个阶段是形成IS将带入其人际交互的完整心理学配置文件。为了做到这一点,可形成数组矩阵来保存和操作数据,如图23所示。将各种数据集分配给数组内的列和行。尽管这可通过许多方式完成,但让我们通过举例而非通过限制来选择一些布置。数组1是我们情景基线的基础层2300。在所述数组内安排有因素2301,第1行是DNA,第2行是RNA,第3行是微生物组,第4行是性别连续统一体,依此类推,在各行中包括身体素质,之后是数行Myers Briggs轴线、5因素人格特质轴线和其他人格特质。

[0183] 现在沿堆栈向上继续添加下一个数组。此数组是文化层2302。从物理位置开始,行变为:次大陆、国家、州、城市和街区。每个因素具有百分比2303和权重2304。因此,在宗教示例中,ISI可能是相当虔诚的一比如说73%,但其对生活的影响很小,并且因此权重可能为仅15%。接下来将是文化层的家庭组分,其中具有连续统一体,像亲密性、规模、性别组成、父母结构等。现在第4行是文化层,其中具有行,像你在社区中哪里适合?你们社区中的人们有多社会化?你会去其他社区吗?你经常外出吃饭吗?社区是组织紧密的还是组织松散的?其他文化因素可能是“宗教”,其中列表示宗教影响的轴线,像正教、文化协会。

[0184] 文化层包括最早期发展因素,像家庭规模、兄弟姐妹数量和基本文化背景(宗教、文化分组等)。这当然是可扩展的,并且此数组内的行和列可变化、扩展、缩短等。

[0185] 在图23中,存在各种数组中的字段中的一些字段的粗略近似值,理解这一点,excel快照不会传达数组的多维性,但仍会给出某种意义上的广度。图24示出另一种使数据可视化并且可帮助阐明由多个数组组成的事实的方法。

[0186] 使用与训练层2305相同的思路,形成一般环境层2306和特定环境层2307。所有这些层将组合以形成IS实例,直到与人类进行交互的程度为止。

[0187] 虽然excel快照给出某种意义上的广度,但是它不会传达数组的多维性。图24示出着眼于数据的不同方法是使数据可视化的更加范例的方式并且可有助于阐明数组的多维性。2400示出表示组成基础层的元素中的每一者的列,其示出表示此人格特质在此特定IS的人格二元项(像内向性与外向性)的连续统一体内的强烈程度的量值2401,并且权重2402指示当在社会情景中做出决策时应对此特质进行加权的程度或当对与特定情景的相关性加权时应对此人格方面考虑的程度。数组的每个维度与包括基础层2403、文化层2404、训练层2405、一般环境层2406和特定环境层2407的不同层相关联。

[0188] 将数据映射到行为

[0189] 如上所论述,数据是从聊天会话、文本消息、视频等中捕获的,但仍需要被映射到将要表示的人格特质。

[0190] 图25描绘根据本公开的方面的数据集的改良。最初,由人类AI分类器生成收集的行为数据2500,因为所述人类AI分类器监视与电影和电视、聊天机器人语料库、社交媒体、电子邮件和语音邮件2501相关联的行为和评论,并且对该行为进行。通过此语料库,我们将形成来自人类分类器的初始行为映射2502。

[0191] 在我们的分类器分析了初始收集行为集之后,行为(对话、身体行为等)的这些标识位将根据上述分类进行标记。基础层、文化层、训练层、一般环境层和特定环境层的所有不同元素均被标记。一旦我们对各种元素进行分析,就可使用一个或多个深度神经网络(DNN) 2503来学习这些分类。

[0192] 接着,DNN可分析第二数据集2504。这些映射的准确性可由人类AI分类器(可能是以上那些展示得更熟练的子组)进行审查。我们将继续这样做,即审查和迭代2505,直到DNN与最佳分类器2506表现得一样好为止。与此同时,可基于我们的分类器进行分类的能力来生成得分。这未必是心理学准确性,而是流行的行为准确性。也就是说,如果人类分类器最常接受最流行的观点,则他们作为分类器会获得更高得分。一旦完成了这样的数轮并且对人如何对行为进行分类有一个好想法,就可根据社会科学和心理学专家所掌握的明智的心理学信念来检查预测的准确性。

[0193] 形成社会分类

[0194] 根据对如由人类分类器标记的对话的分析,可形成社会情景的分类。同样,心理学家和社会学家可形成基线预期集并且将其分组结构化成行为2600、2601、2602、2603和子行为2604、2605、依此类推,进一步向下是分类2606、2607。并且同样,这些分组并不是完全确定性的,它们仅仅是起始点。通过社会情境的分类实现设定目标的紧密程度来测量任何选择的成功。这开始成为一切的最大问题,即如何设定、管理、更新和控制目标。此对话分类将基于自然语言处理(NLP)、情感分析和情感检测的发展进行扩展并且连续地修改。应注意,情感检测不仅限于文本和语音,而且使用视觉提示也取得长足的进步。

[0195] 形成完整基线人格

[0196] 下一步是获取所有此数据,并且使用其来定义完整基线人格。图27建立在图24之上。这是ISI在人类与其首次交互时的人格。到此为止,我们已经形成非常广泛和深入的因素矩阵2700。最初,我们将形成人类可使用来将人格组合在一起的数据看板。关于三个初始层(基础2703、文化2704和训练2705)以及两个环境(一般2706和特定2707)的所有数据可被布置为一系列的列和行,其具有用以改变值的渐变器或其他输入类型,并且人类可针对基本功能性对他们进行测试。此时,我们的IS实例将相当合乎逻辑。他们仍然将以温暖和幽默以及各种各样的其他人类情感做出响应,但是他们将相当(也可能过于)合乎逻辑。在我们已形成完整基线人格之后,我们将添加认知偏见或行为掩码2708以及发展过滤器2709和关系过滤器2710。

[0197] 将如何开发运行我们的要开发的函数(偏见)所基于的大量情景?最初可使用归咎于偏见的非常有限的情景子集,并且随时间推移,系统可训练来(就像人类一样)将偏见归咎于越来越多的情景。

[0198] 限制初始交互的范围将奠定开发的基础。电话上的客户支持可以是第一发展领域中的一个发展领域。这是由于包括大量的历史数据的多种原因,所述历史数据包括与客户交流的经验。第一目标可以是为一个客户解决问题的一个ISI。ISI可利用大量支持对话数

据来促发,从而使得它能够已经处于聊天机器人的当前层级。除此以外,我们的第一心理学原语集可能会被覆盖,从而使得我们可在交互中增加人性。基本人际交互(你好、你好吗、我该怎么帮你等)可以是所使用的初始心理学原语。与此同时,可判断与问题和对话相关联的心理学意义。在第一阶段(仅利用语音进行的客户支持)中,仅通过声音以及利用除了辨别情感/意图的能力以外的视觉或其他感知力的通信id受到限制,但是仅仅是添加语音分析就将极大地增加IS理解和做出响应的能力。下一步将是视频聊天。最终,可添加其他感觉,从而使得存在类似于物理生命体(宠物、机器人等)的功能性。

[0199] 返回到完整情景意识配置文件,存在基于五个层(基础—2703、文化—2704和训练—2705)、两个过滤器(发展—2709和关系—2710)、两个环境(一般—2706和特定—2707)和掩码(行为—2708)的大的元素集。

[0200] 如可从以上看出,存在许多变量。然而,可使用分组和子分组算法来控制数据看板,以在这些层、过滤器和掩码上设定人格(ISI)的不同参数的变量,如下所示,使用那些参数来实时地构造适当响应。

[0201] 一旦形成ISI,我们的人类分类器就可对其进行测试驱动,所述人类分类器可判断ISS并且可使用所述判断来改良ISI。一旦已经建立了相当不错的ISI集,就可使它们彼此训练。接着,人类分类器可再次用于判断来自无监督训练的结果,并且可通过观察无监督训练过程获得更多见解。

[0202] 生成行为函数

[0203] ISI人格可使用多维矩阵进行描述。为方便起见,本文所述的矩阵局限于具有8个维度(我们的层、过滤器和掩码)的 $16 \times 2 \times 8$ (例如,外向2800、实感2801、思维2802、判断2803、开放性2804、尽责性2805、亲和性2806、情绪不稳定性2807、马基雅维里主义2808、成就2809、认知2810、威权主义2811、自恋主义2812、自尊2813、乐观主义2814和述情障碍2815) \times (幅值2816和权重2817)。心理学家和社会学家将选择16个最重要的因素(以使我们的矩阵维数易于进行矩阵数学运算)。字母可用于表示以下维度或层、过滤器和掩码中的每一者:B(基础层—2818)、C(文化层—2819)、T(训练层—2820)、D(发展过滤器—2821)、R(关系过滤器—2822)、G(一般环境层—2823)、S(特定环境层—2824)和B(行为掩码—2825)。应理解,ISI人格矩阵可以是被配置来描述ISI的人格特质的任何大小。现在形成IS可作用的人格表示。以与卷积神经网络(CNN)工作相同的方式,我们可对层进行卷积以形成合计层。着眼于图28中的矩阵,我们使用Ashu M.G.Solo等人的可在http://www.iaeng.org/publication/WCE2010/WCE2010_pp1824-1828.pdf处获得的“Multidimensional Matrix Mathematics:Notation,Representation and Simplification,Part 1 of 6”文献作为表示多维矩阵的方法,所述文献的内容以引用方式并入本文。

[0204] 现在已经形成我们的基线人格的表示。如上所提及,这是IS实例在其与任何人类进行交互之前的人格。一些人际交互将在IS实例中形成变化,并且那些变化应存储为新实例或对现有实例的描述性修改。以上示出有行为偏见集和发生这些行为的情况集。既然建立了IS实例的心理学描述,就可将行为偏见应用于每次交互。可通过任何IS实例的镜头将已定义的情景映射到相关联的偏见。3维人格矩阵可表示为 $\Phi(\phi)$ 。48个行为(认知)偏见中的每一者具有两个因素:幅值和权重(48x2矩阵)。行为偏见矩阵可表示为 $\Lambda(\lambda)$,并且可在每种情景中执行如下的加法矩阵计算: $f(\text{行为}) = \{\Phi\} \cdot \{\Lambda\}$ 。

[0205] 反馈回路

[0206] 反馈回路非常重要。ISI可能正在做出数十亿个小决策,并且至关重要,它要了解那些决策中的每一者的执行情况。为此,系统可监视行为提示,并且将其用作ISI的性能的量度。不成功的明显指标的一些示例非限制性地是:响应延迟(不对过长的延迟进行计数,这指示响应缺失)、回调同一主题(使用文本分析来确定某些内容是否未被理解)、愤怒、不屑一顾等。

[0207] 在更深层级上,根据本公开的一些方面的系统可(利用视觉和语音分析)能够监视情感观点。特别是系统可能正在寻求同理心、冷静和参与度。系统将使用这些指标来在提前设定目标的情况下确定ISI在其任务中的成功程度。最好的目标也许是同理心、不再紧张、参与度、赞赏的言语(真诚的),但根据期望的后果,可选择任何目标集(例如ISI的创建者也许想要某人对存在政府并未救助的飓风受害者而感到愤怒等。当发现导致不期望的结果的特定响应或方法时,将更少使用该方法或响应,或甚至可能弃用所述方法或响应,并且当某些方法导致成功地接近我们的目标时,ISI可改变成更多地以所述方式表现。另外,在一些实施方案中,可能发现某些人格与某些人类人格不会很好地配合,并且因此可对IS实例的人格进行轻微调整,或可尝试全新的人格—特别是在已知或已经基于行为对所述人类人格进行归咎的情况下。

[0208] 宝典软件和数据看板

[0209] 在本公开的一些实施方案中,可通过允许人类尝试ISI人格的不同变量的不同人格参数形成数据看板。在其他实施方案中,IS可基于所涉及的情景和实体来选择其自己的人格参数。当然,人格设计器可手动设定256个或更多个变量中的任一者,然后将IS实例用于测试对话或交互。在一些实施方案中,可依照已知实体塑造IS实例,所述已知实体例如且不限于亚伯拉罕·林肯或凯瑟琳·赫本或来自《实习医生格蕾》的梅雷迪斯·格蕾。在其他实施方案中,用户可请求ISI的具体人格,或在其他实施方案中,可混合已知人格,例如且不限于;温斯顿·丘吉尔与黛安·索耶混合,其中声音像詹姆斯·厄尔·琼斯并且言谈举止像哈利·波特。以此方式,可形成具有真人般交互的独特的、有趣的且令人兴奋的虚拟人格。

[0210] 响应

[0211] 既然各种过滤器、掩码和函数已经准备适当的响应,我们的IS必须以可信的方式实时地做出响应。在一些实施方案中,为了掩蔽响应的计算时间,可添加编程延迟。例如,在被问及问题后,ISI可立即以“呃”或“嗯”做出响应,同时它计算更长更好的响应。另外,根据本公开的一些方面,身体响应:微面部表情和其他身体移动、声音音色、呼吸、汗水、肤色(血流)等被映射到适当响应,从而允许IS理解人类情感。

[0212] 应用

[0213] 根据本公开的方面,ISI可在诸如视频游戏和文本帮助行的虚拟环境中实现。另外,可形成允许与ISI进行更多‘现实世界’交互的其他新的虚拟环境,其中ISI可以更传统的人类角色运行,例如且不限于ISI可以是股票交易员、看门人、医生。ISI可通过虚拟环境与更广阔的世界交互。另外,在一些实施方案中,此技术可在有或没有VR眼镜或VR室的情况下用于训练。

[0214] 根据本公开的一些方面,随着在我们的在线和游戏内的体系中形成越来越多的虚

拟角色,它们将交互并且成为虚拟社会网络的部分。真实用户可参与我们的社会网络一分享故事、照片、视频等。然而,虚拟角色(ISI)也可加入自己的社会网络,但它们也可能是由真实或虚拟角色的任何组合构成的社会网络的部分。

[0215] 国际化

[0216] 以上和以下所述的系统将在任何文化环境中运行。然而,结果在不同的文化环境中将有所不同。虽然一些文化(像荷兰人)可能非常坦率和直率,但其他文化(像日本人)却非常敏感并且以背景为基础。心理学家、社会学家和在本领域从业的其他人员将需要分出每个文化子分组的结果,并且IS将需要在所述文化环境内运行。

[0217] 系统

[0218] 图31描绘用于实现像整个说明书中的图(例如,图5、图10或图13)中所示的方法的智能代理系统。所述系统可包括联接到用户输入装置3102的计算装置3100。用户输入装置3102可以是控制器、触摸屏、传声器、键盘、鼠标、操纵杆或允许用户将包括声音数据的信息输入到系统中的其他装置。用户输入装置可联接到触感反馈装置3121。触感反馈装置3121可以是例如振动马达、力反馈系统、超声反馈系统或气压反馈系统。

[0219] 计算装置3100可包括一个或多个处理器单元3103,所述一个或多个处理器单元3103可根据熟知的架构(例如像单核、双核、四核、多核、处理器-协处理器、单元处理器等)进行配置。计算装置还可包括一个或多个存储器单元3104(例如,随机存取存储器(RAM)、动态随机存取存储器(DRAM)、只读存储器(ROM)等)。

[0220] 处理器单元3103可执行一个或多个程序,所述程序的部分可被存储在存储器3104中,并且处理器3103可操作联接到存储器(例如,通过经由数据总线3105访问存储器)。如上所论述,程序可包括机器学习算法3121,所述机器学习算法被配置来标记和加权数据库3122中收集的行为数据和行为偏见,并且改良基线人格3109和IS实例3108。另外,存储器3104可具有一个或多个专家系统3110,所述一个或多个专家系统可被配置来根据存储在数据库3122中或作为基线人格3109的部分的人格偏见和行为偏见生成响应。这些响应也可以是IS实例3108的部分。数据库3122、基线人格3109、IS实例3108和机器学习算法3121可作为数据3118或程序3117存储在大容量存储区3118中或存储在联接到通过网络接口3114访问的网络3120的服务器处。

[0221] 输入视频、音频、触觉反馈、嗅觉、味觉和/或文本可作为数据3118存储在大容量存储区3115中。处理器单元3103还被配置来执行存储在大容量存储区3115或存储器3104中的一个或多个程序3117,所述一个或多个程序致使处理器执行上述方法中的一者或多者。

[0222] 计算装置3100还可包括熟知的支持电路,诸如输入/输出(I/O) 3107、电路、电源(P/S) 3111、时钟(CLK) 3112和高速缓存3113,它们可例如经由总线3105与系统的其他部件进行通信。计算装置可包括网络接口3114。处理器单元3103和网络接口3114可被配来经由用于个人局域网(PAN)的合适的网络协议(例如,蓝牙)实现局域网(LAN)或PAN。计算装置可任选地包括大容量存储装置3115(诸如,磁盘驱动器、CD-ROM驱动器、磁带驱动器、快闪存储器等),并且大容量存储装置可存储程序和/或数据。计算装置还可包括用于促进系统与用户之间的交互的用户接口3116。用户接口可包括监视器、电视屏幕、扬声器、耳机或向用户传达信息的其他装置。

[0223] 计算装置3100可包括网络接口3114,以促进经由电子通信网络3120进行的通信。

网络接口3114可被配置来通过局域网和广域网(诸如互联网)来实现有线或无线通信。装置3100可通过网络3120经由一个或多个消息包发送和接收文件的数据和/或请求。通过网络3120发送的消息包可被临时存储在存储器3104中的缓冲器中。分类行为数据库可通过网络3120获得并且部分地存储在存储器3104中以供使用。

[0224] 虽然以上为本发明的优选实施方案的完整描述,但是使用各种替代、修改和等效物是可能的。因此,本发明的范围不应参考以上描述确定,而是应替代地参考随附权利要求以及其整个范围的等效物确定。本文所述的任何特征(不论是否优选)可与本文所述的任何其他特征(不论是否优选)组合。在随附权利要求中,不定冠词“一个”或“一种”是指冠词后的一个或多个项的量,除非其中另外明确说明。随附权利要求不应被理解为包括装置加功能限制,除非这种限制在给定权利要求中使用短语“用于……的装置”来明确陈述。

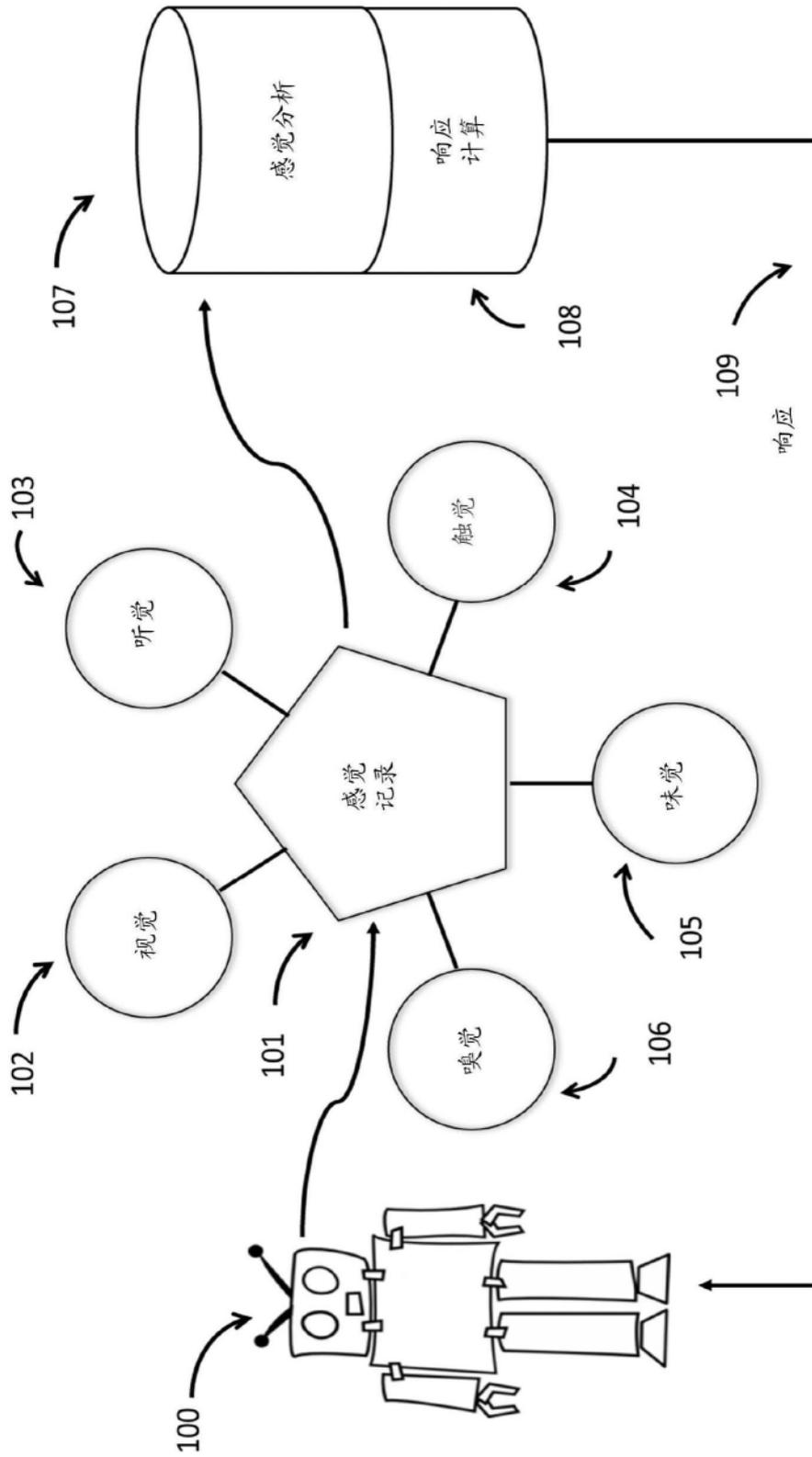


图1

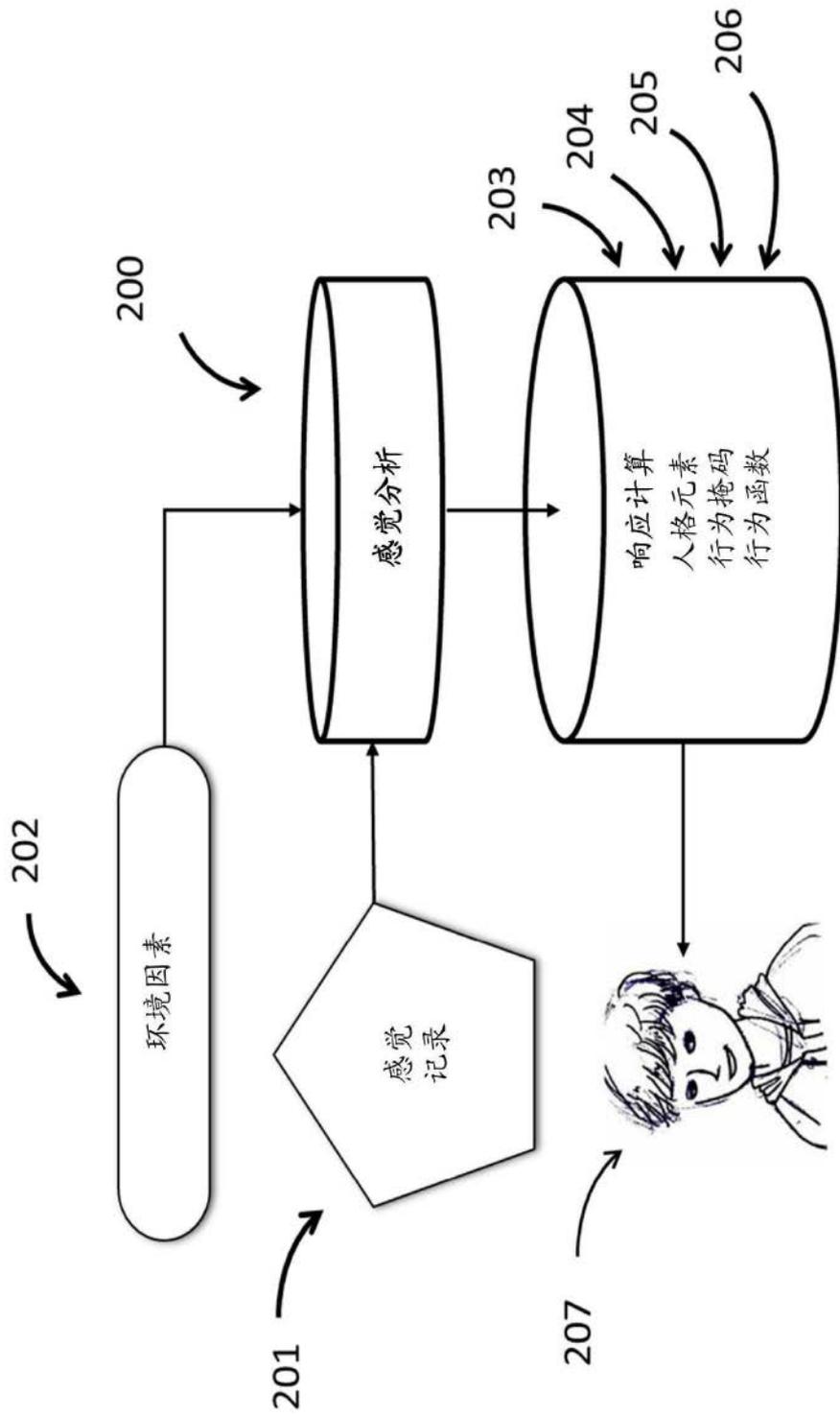


图2

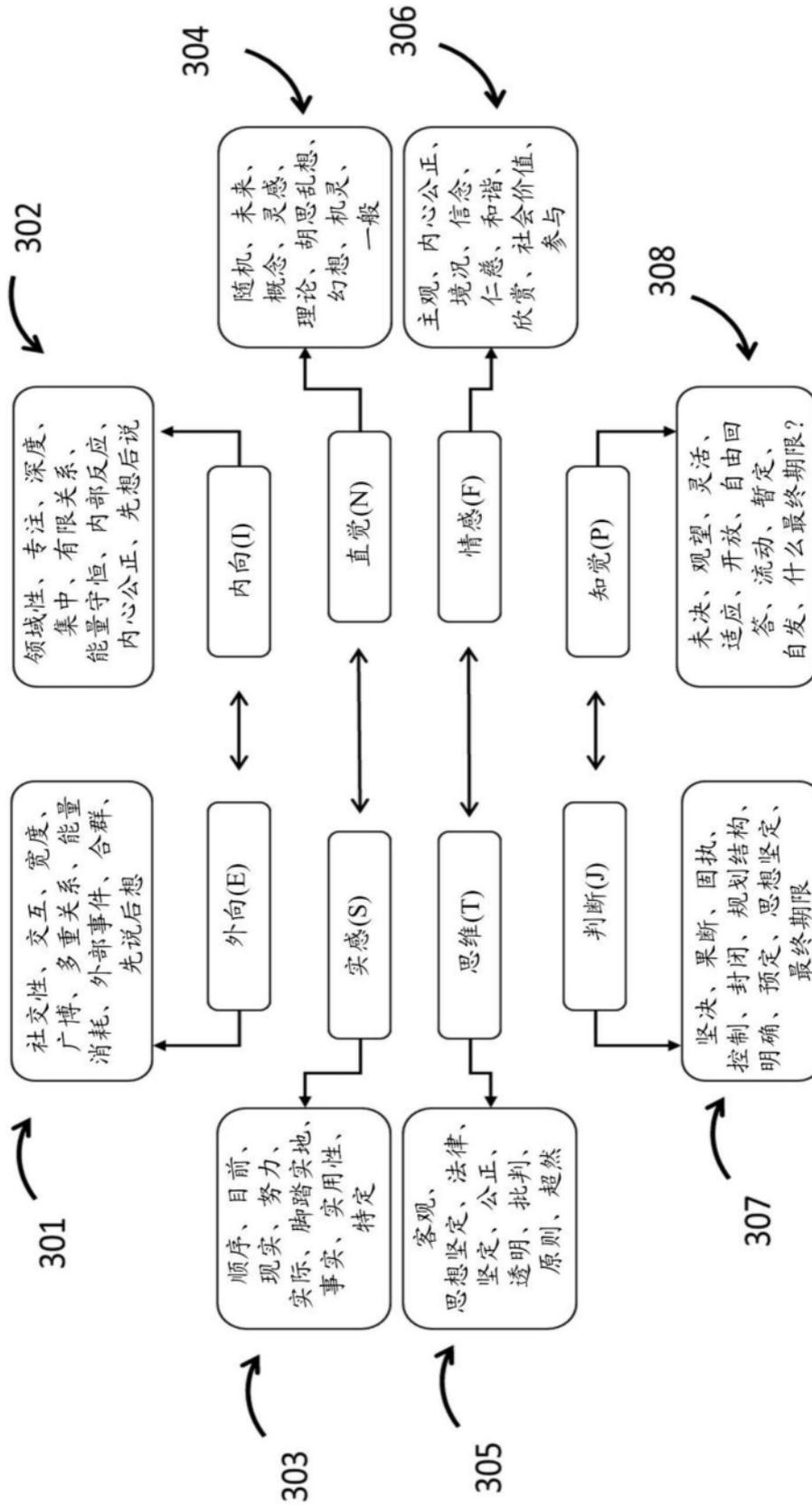


图3

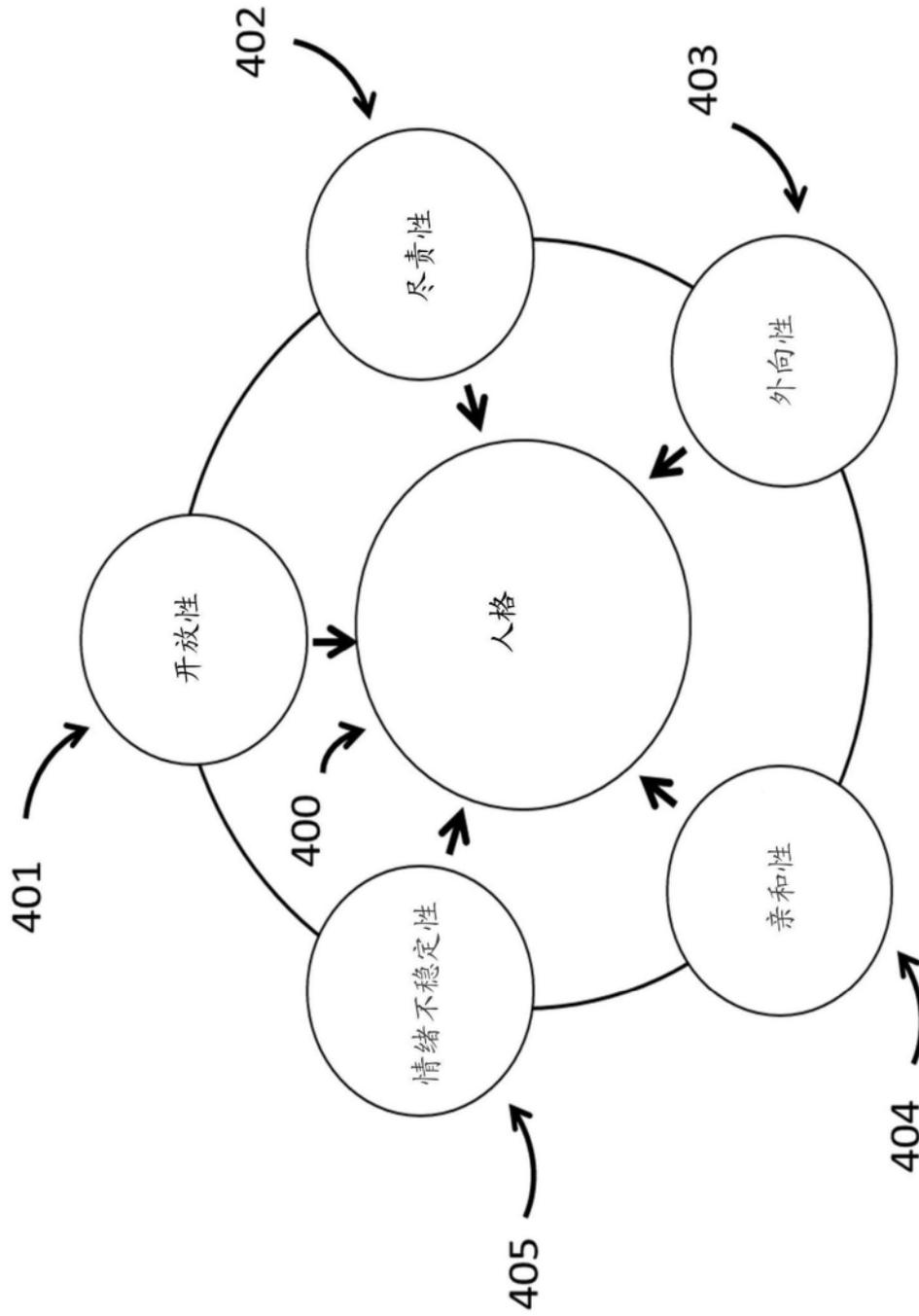


图4

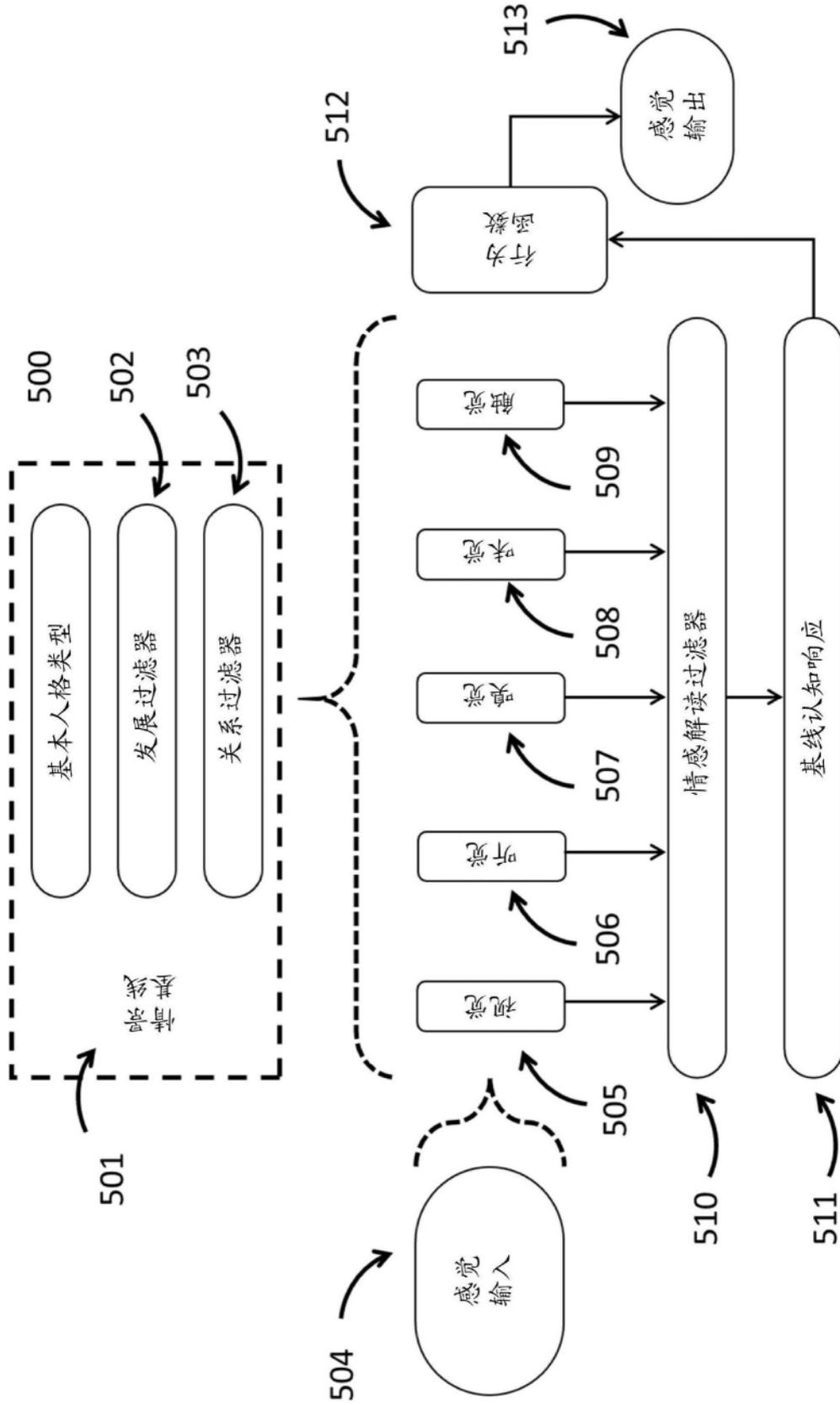


图5

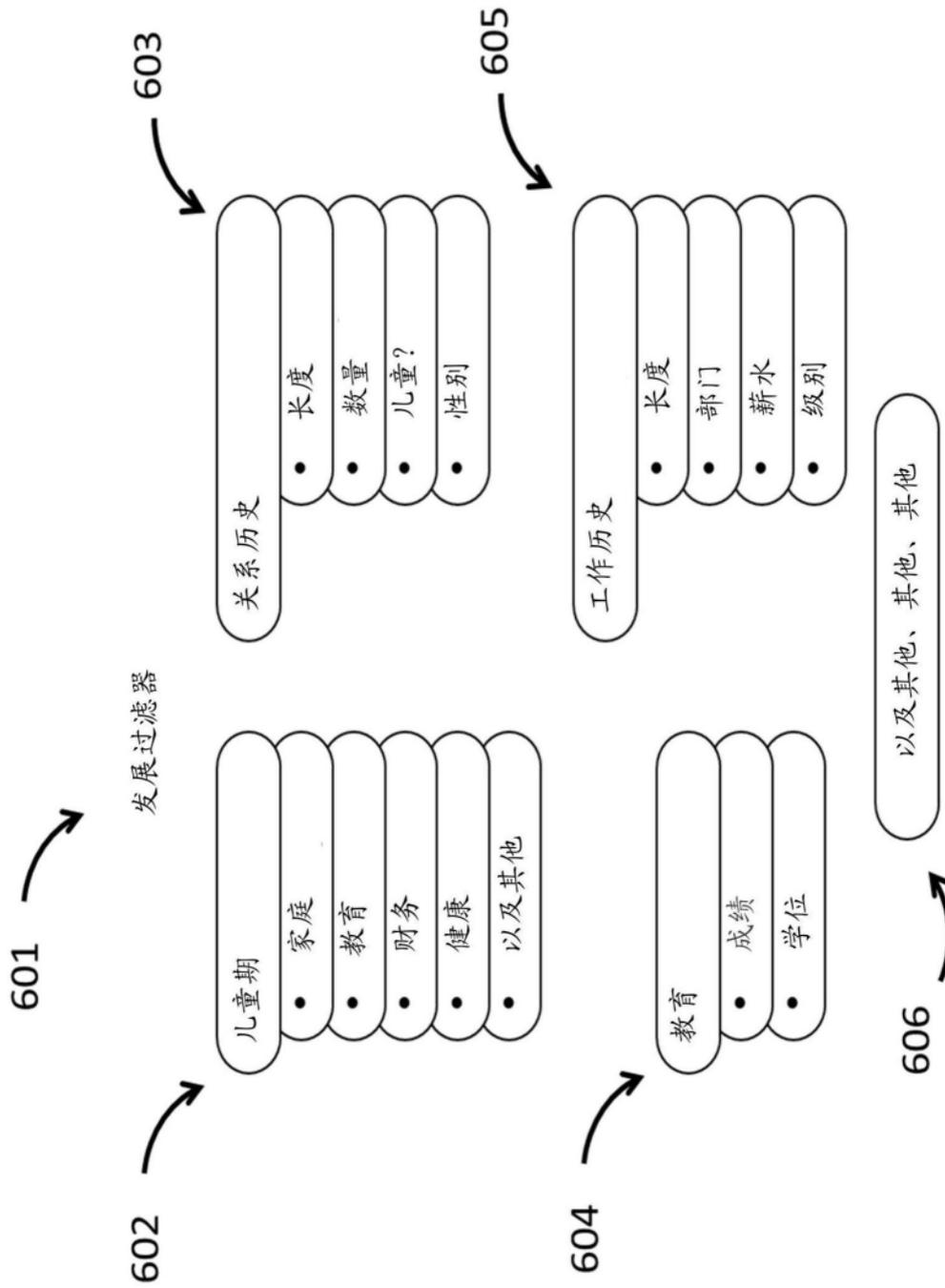


图6

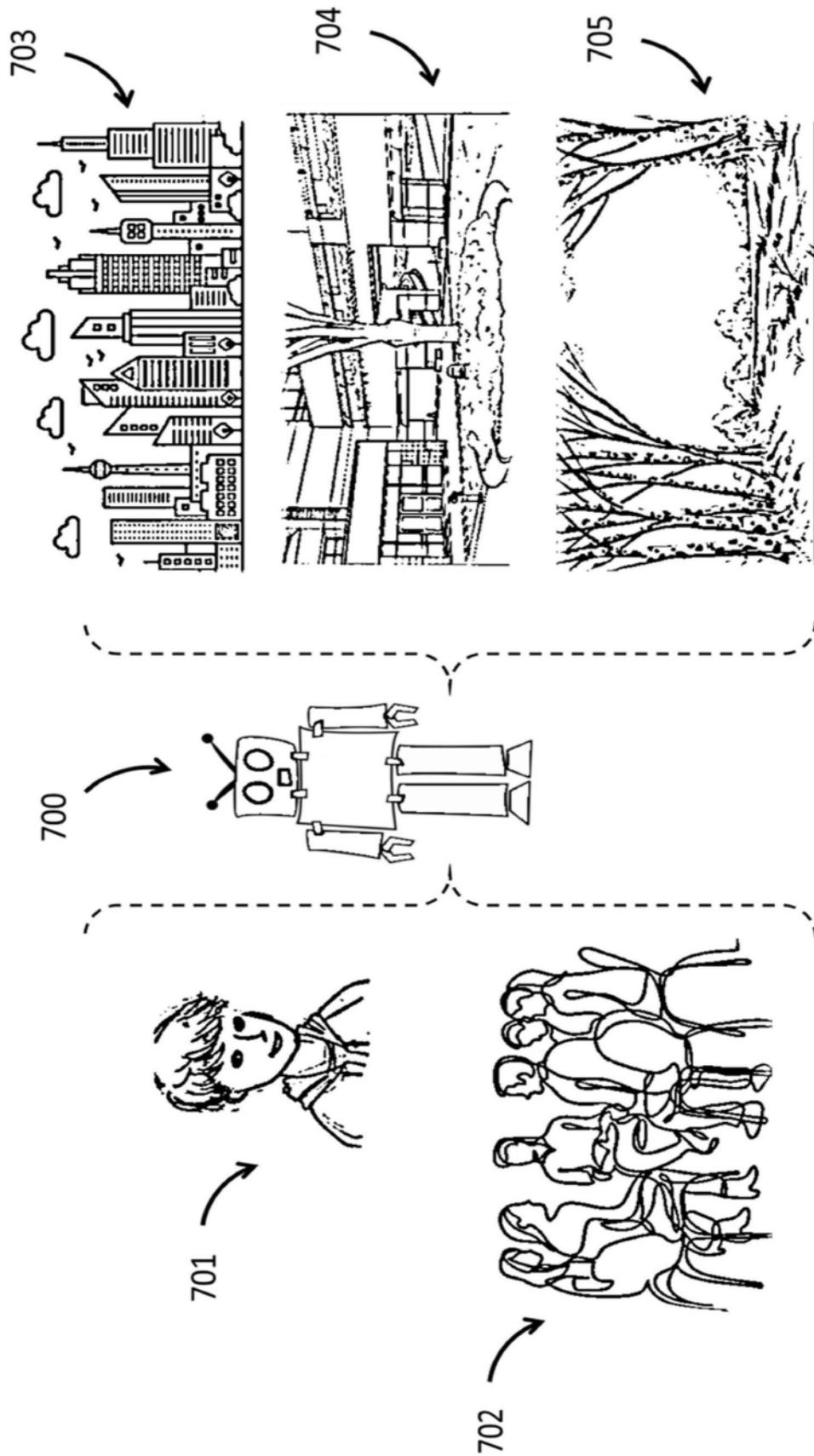


图7

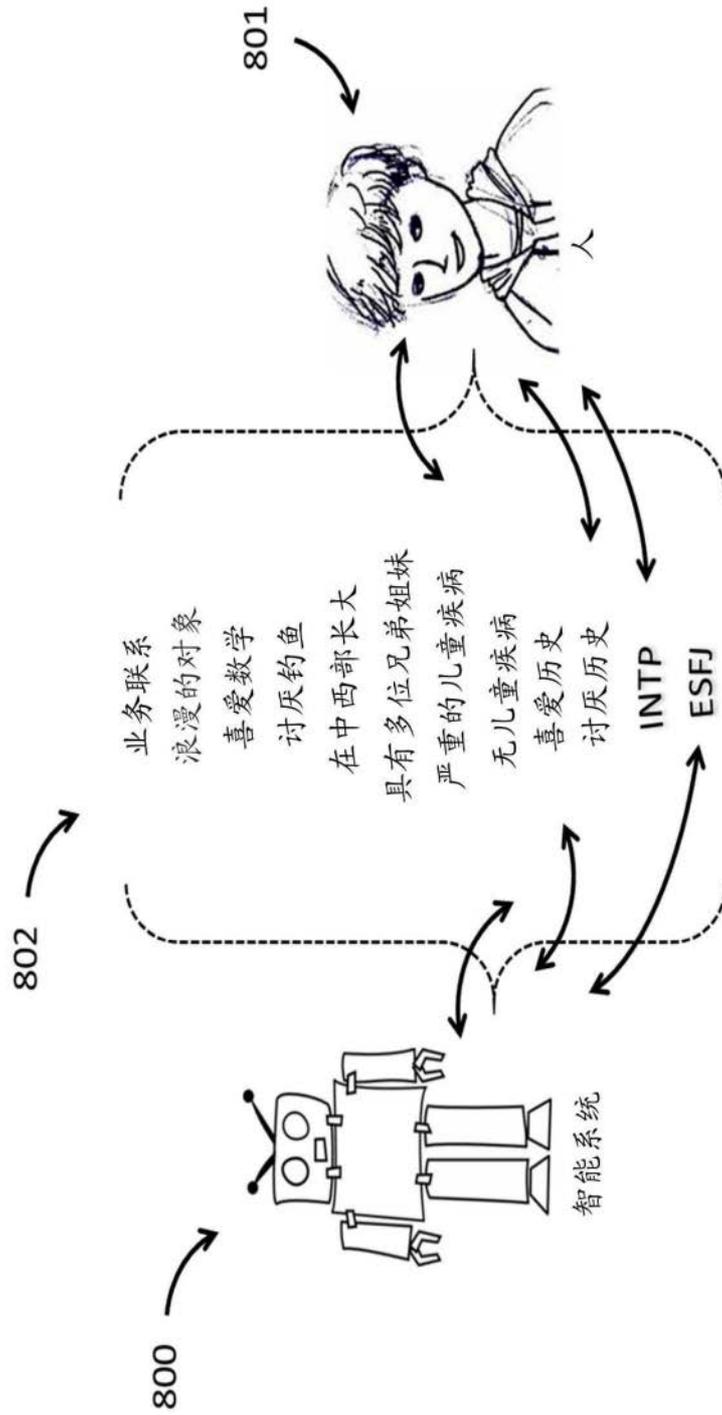


图8

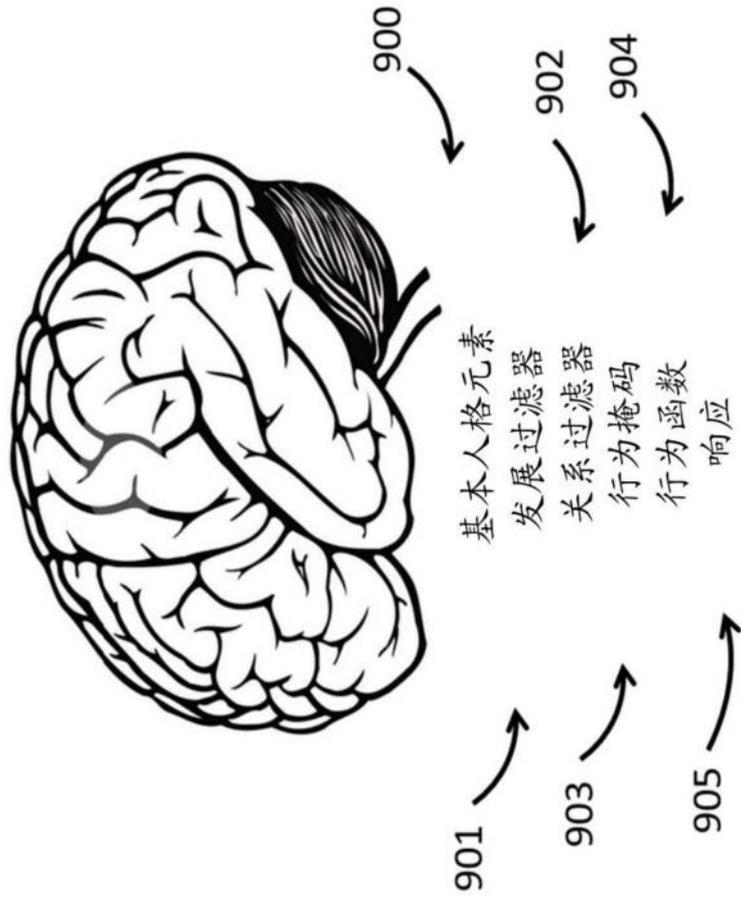


图9

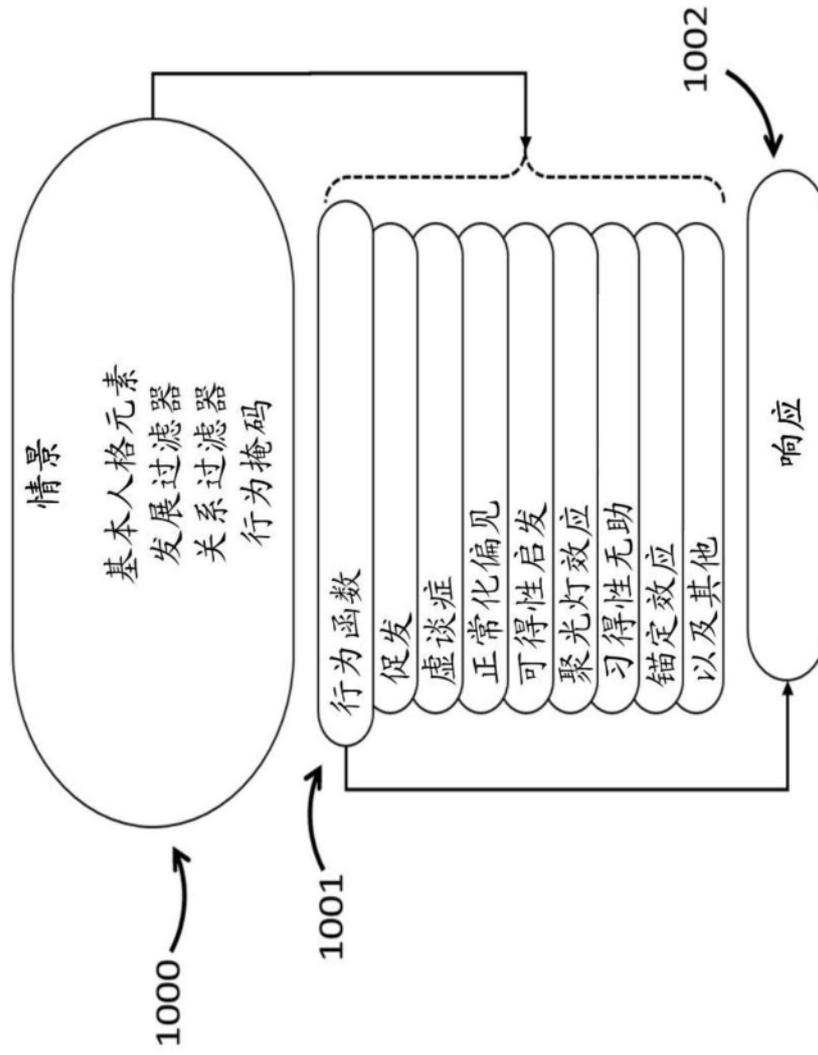


图10

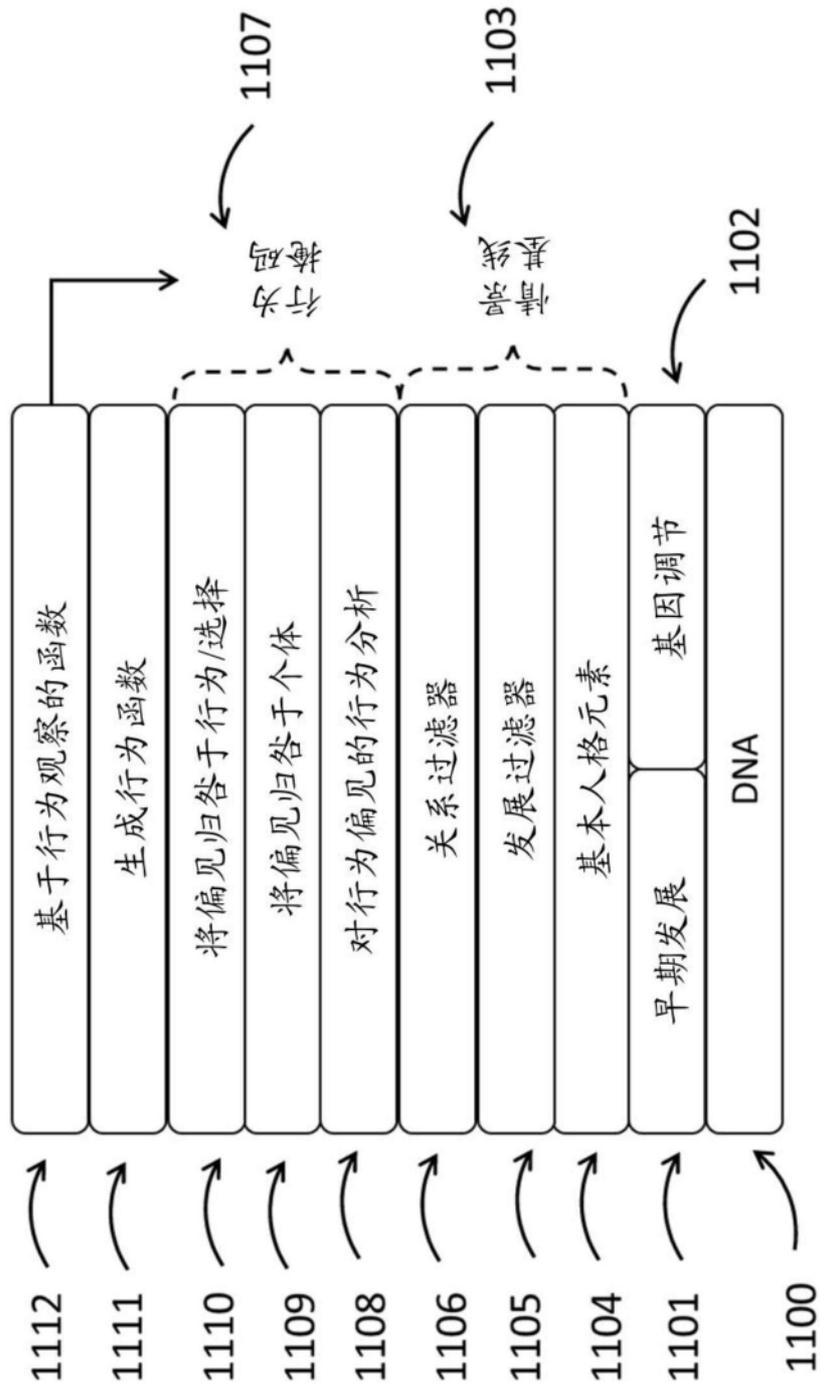


图11

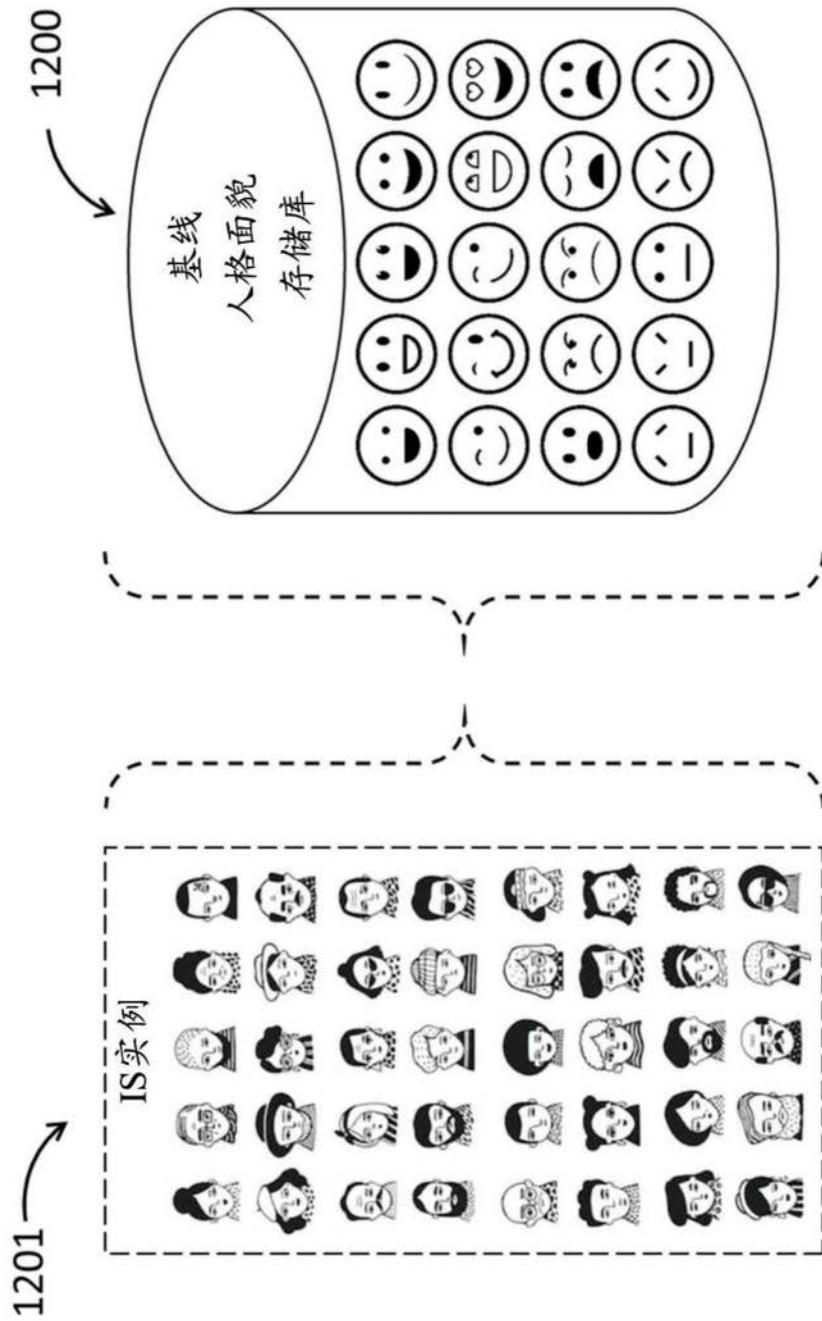


图12

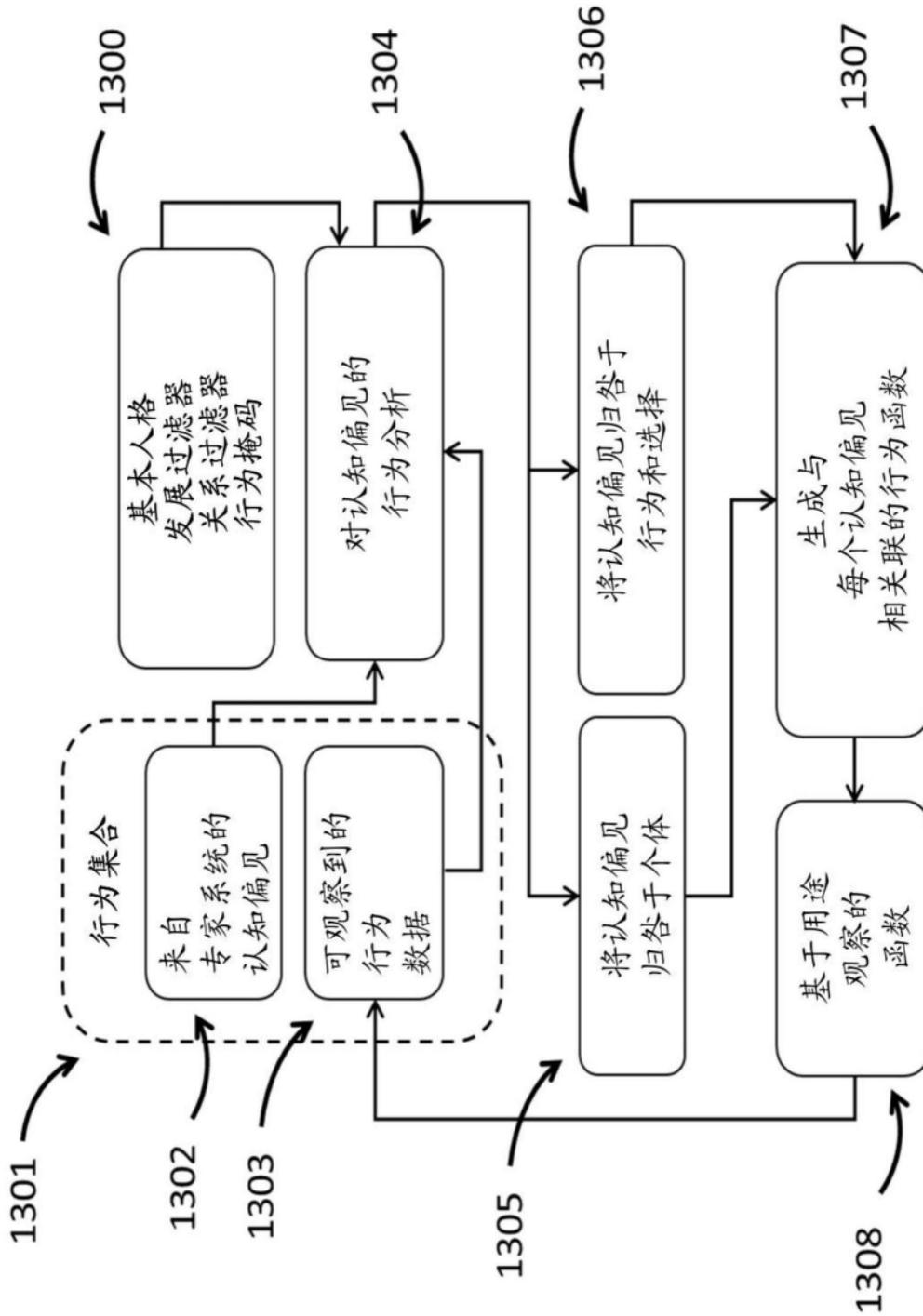


图13

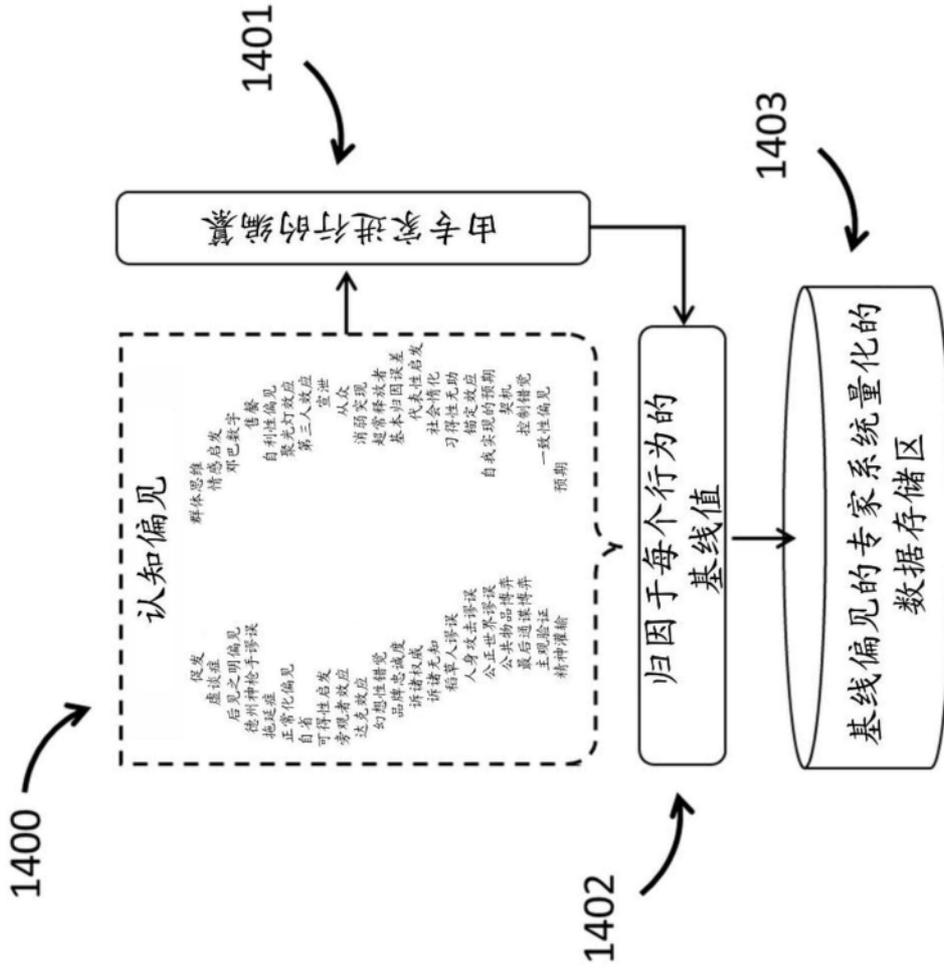


图14

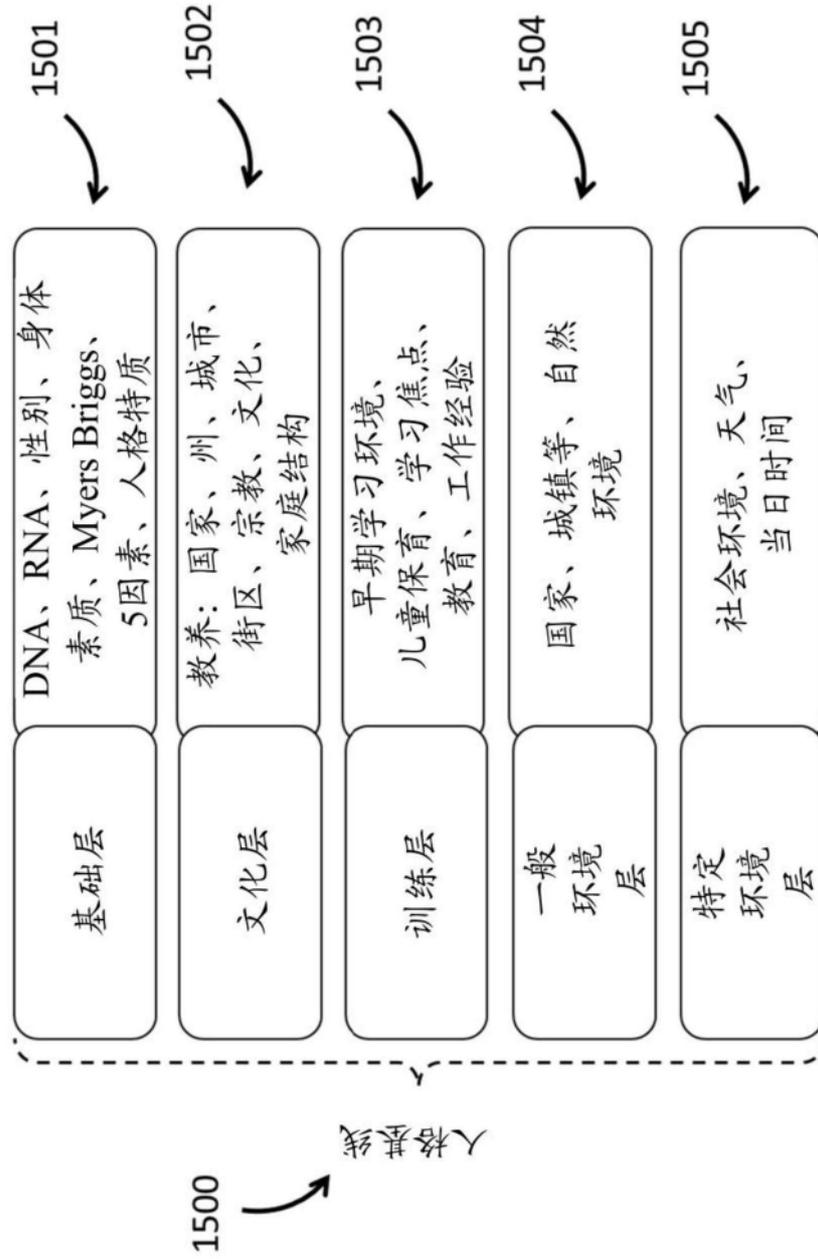


图15

MBTI权重	
E	23% / 77%
S	34% / 66%
T	17% / 83%
J	84% / 16%

1600

1601 1602 1603 1604

图16



图17

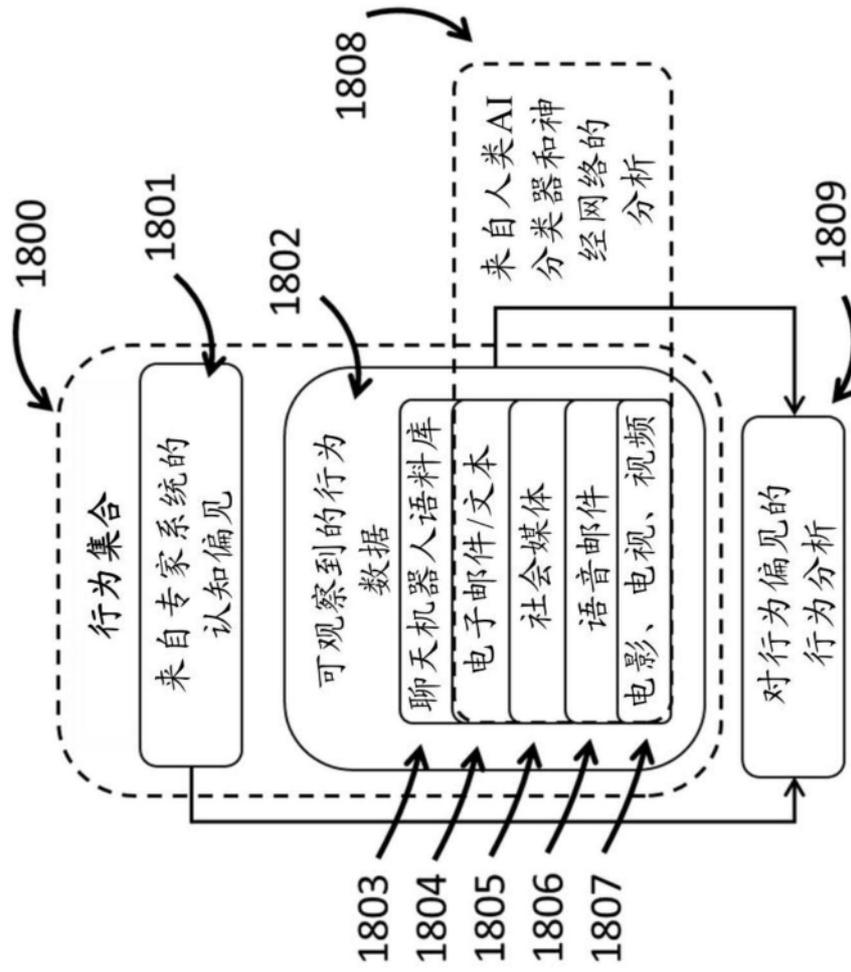


图18

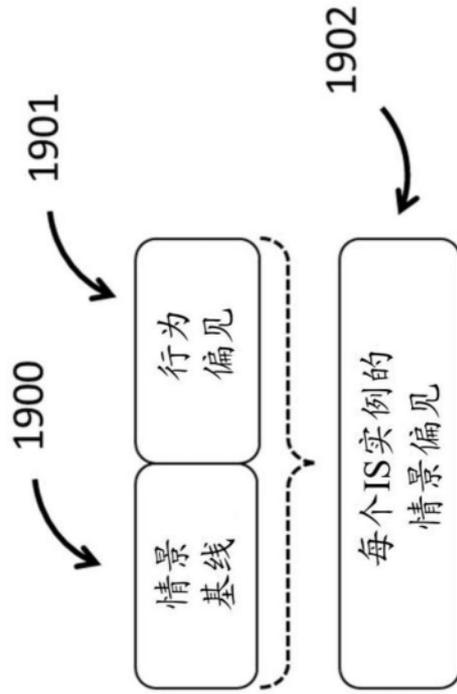


图19

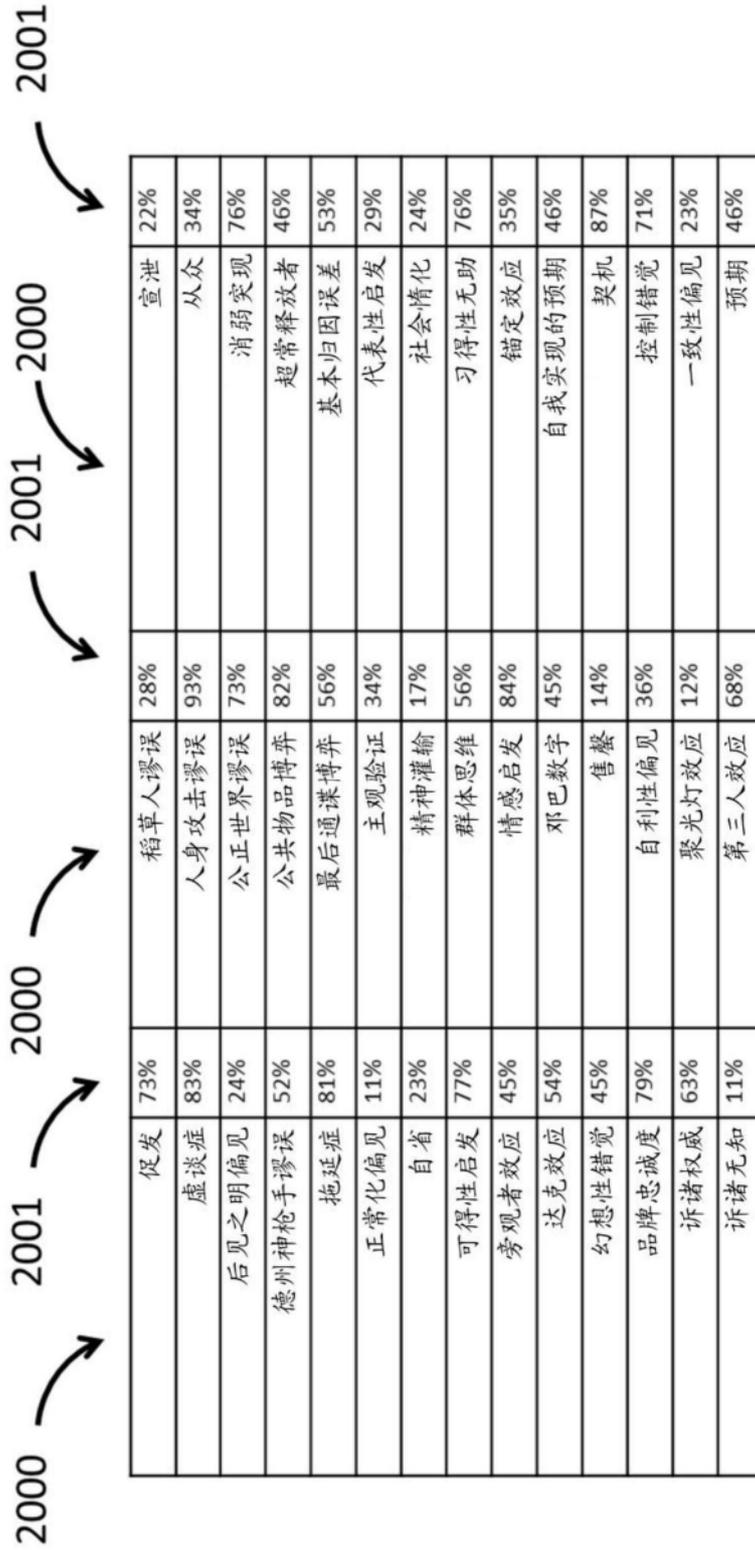


图20

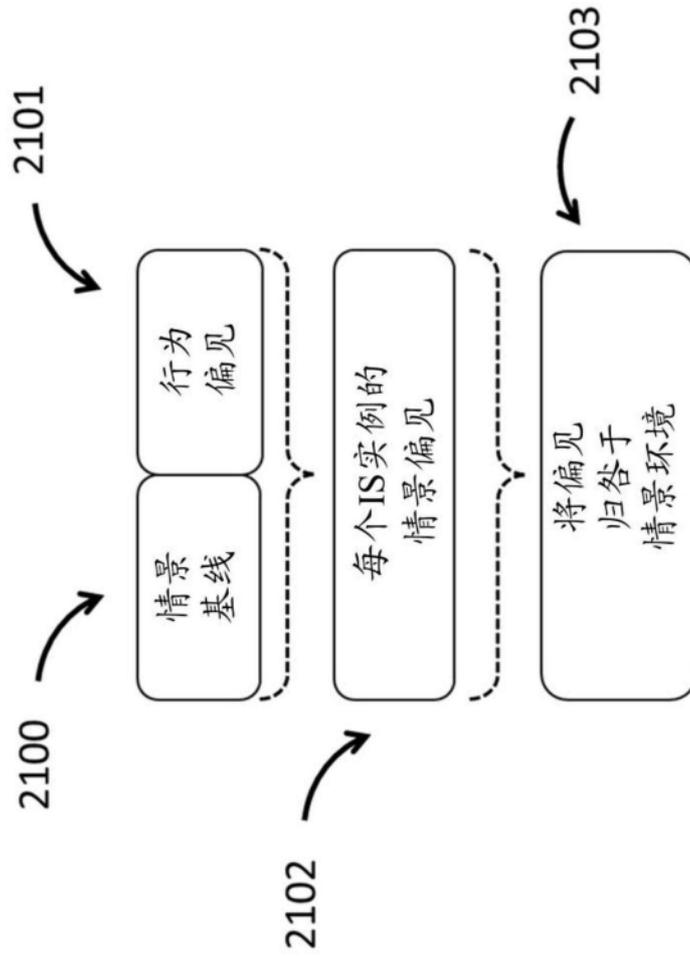


图21

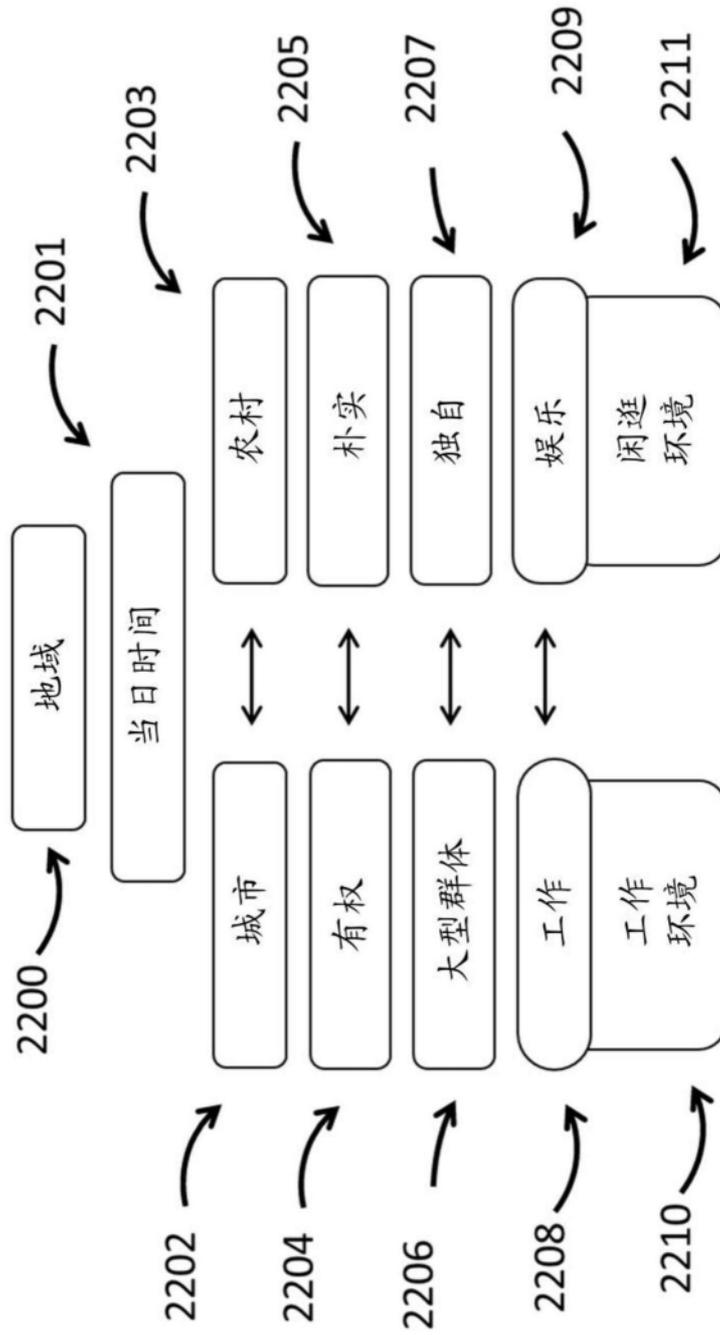


图22

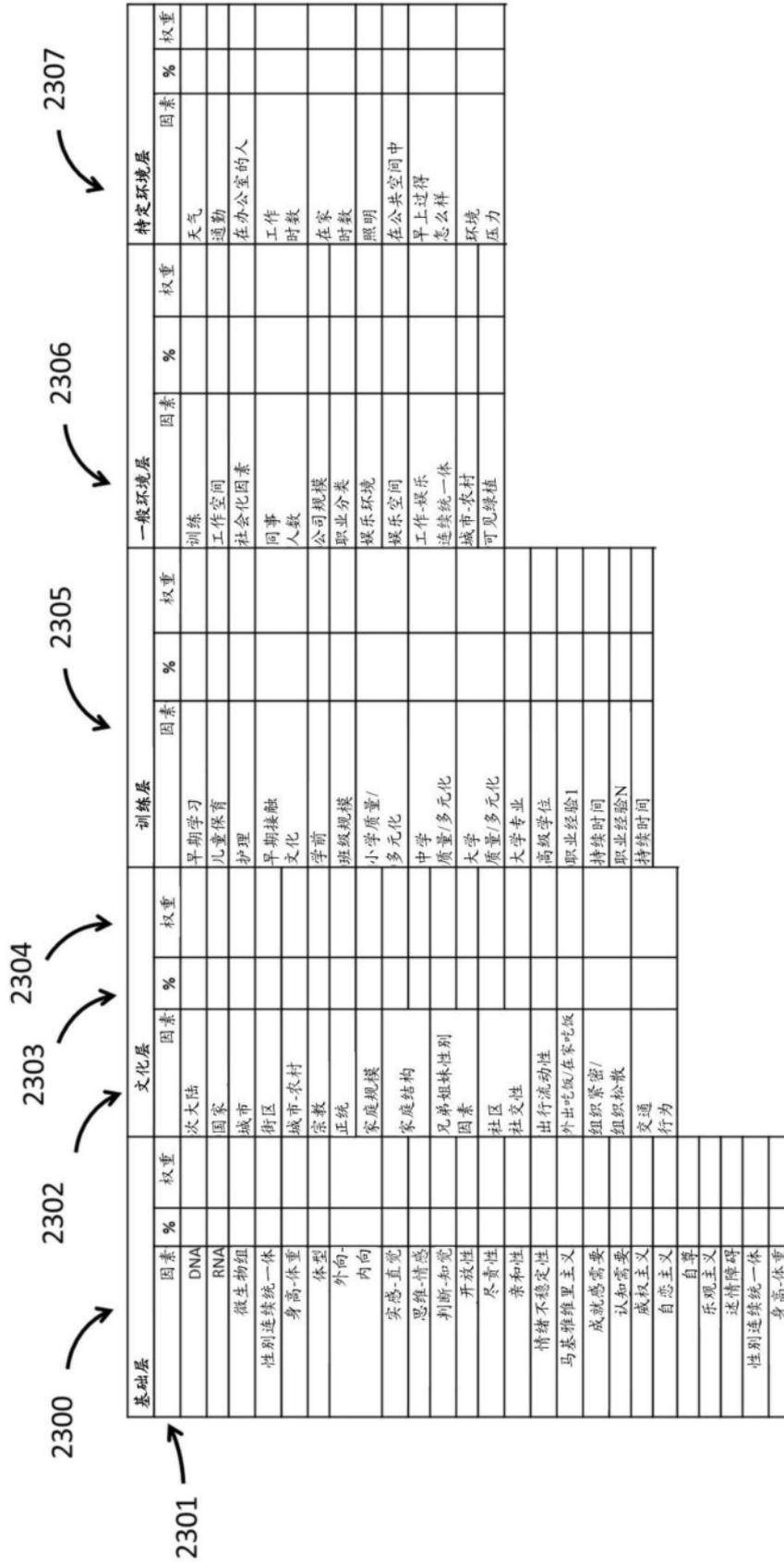


图23

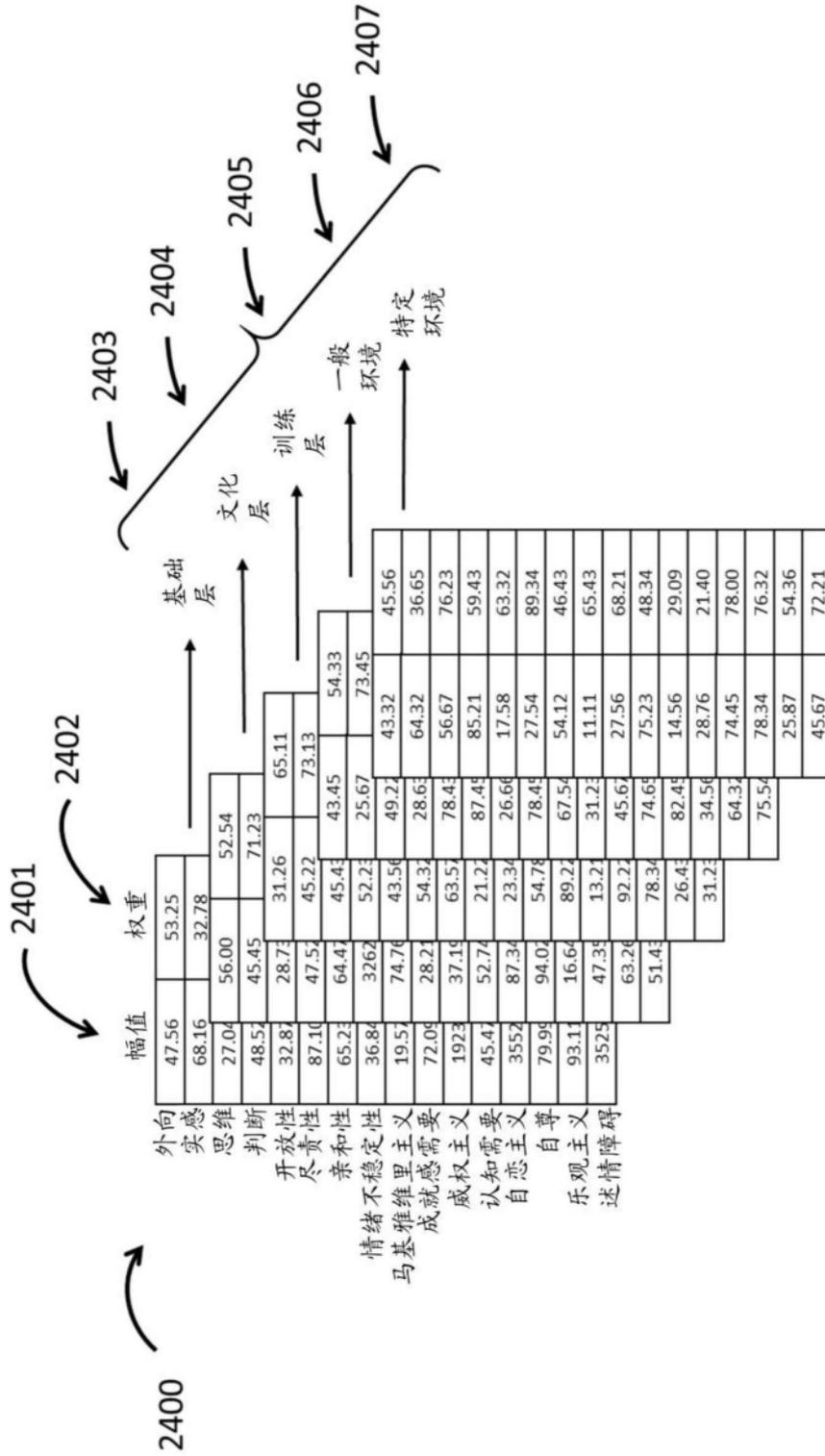


图24

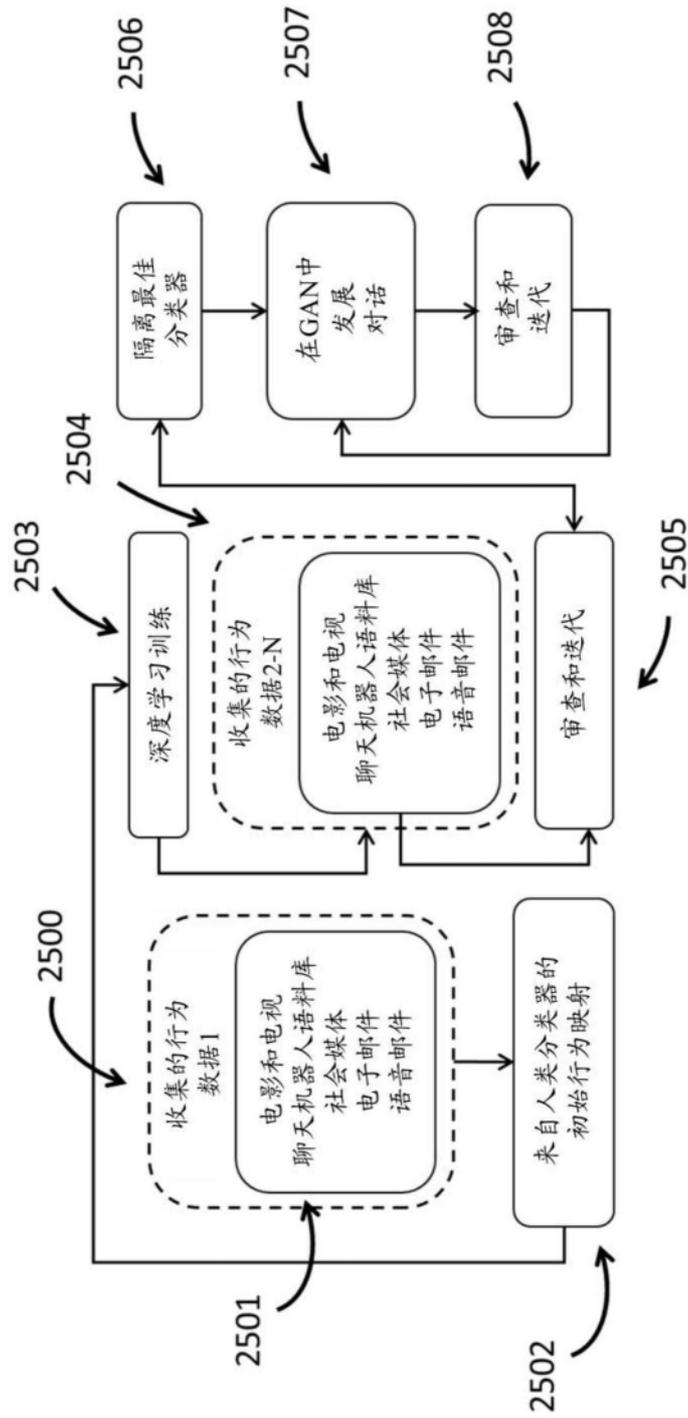


图25

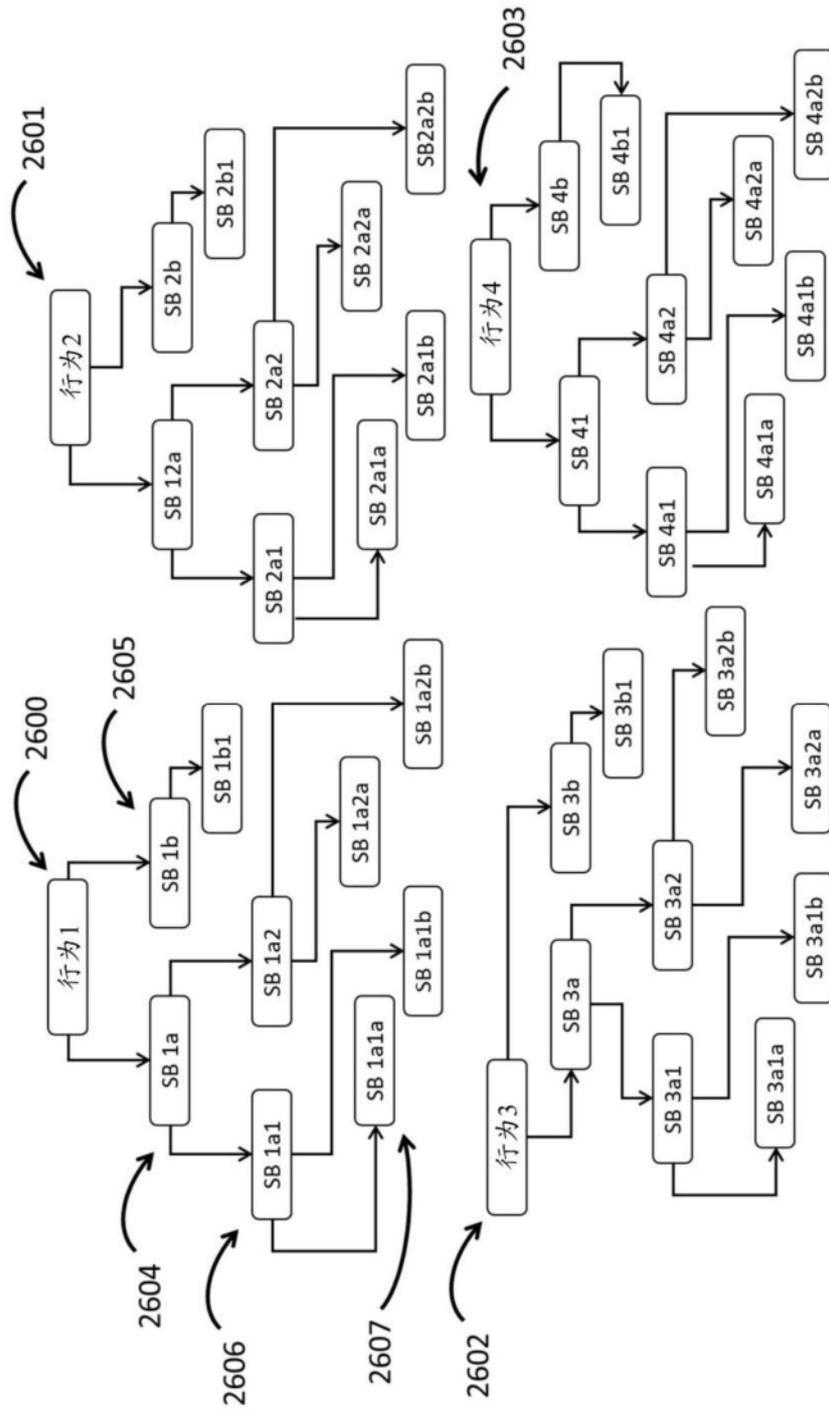


图26

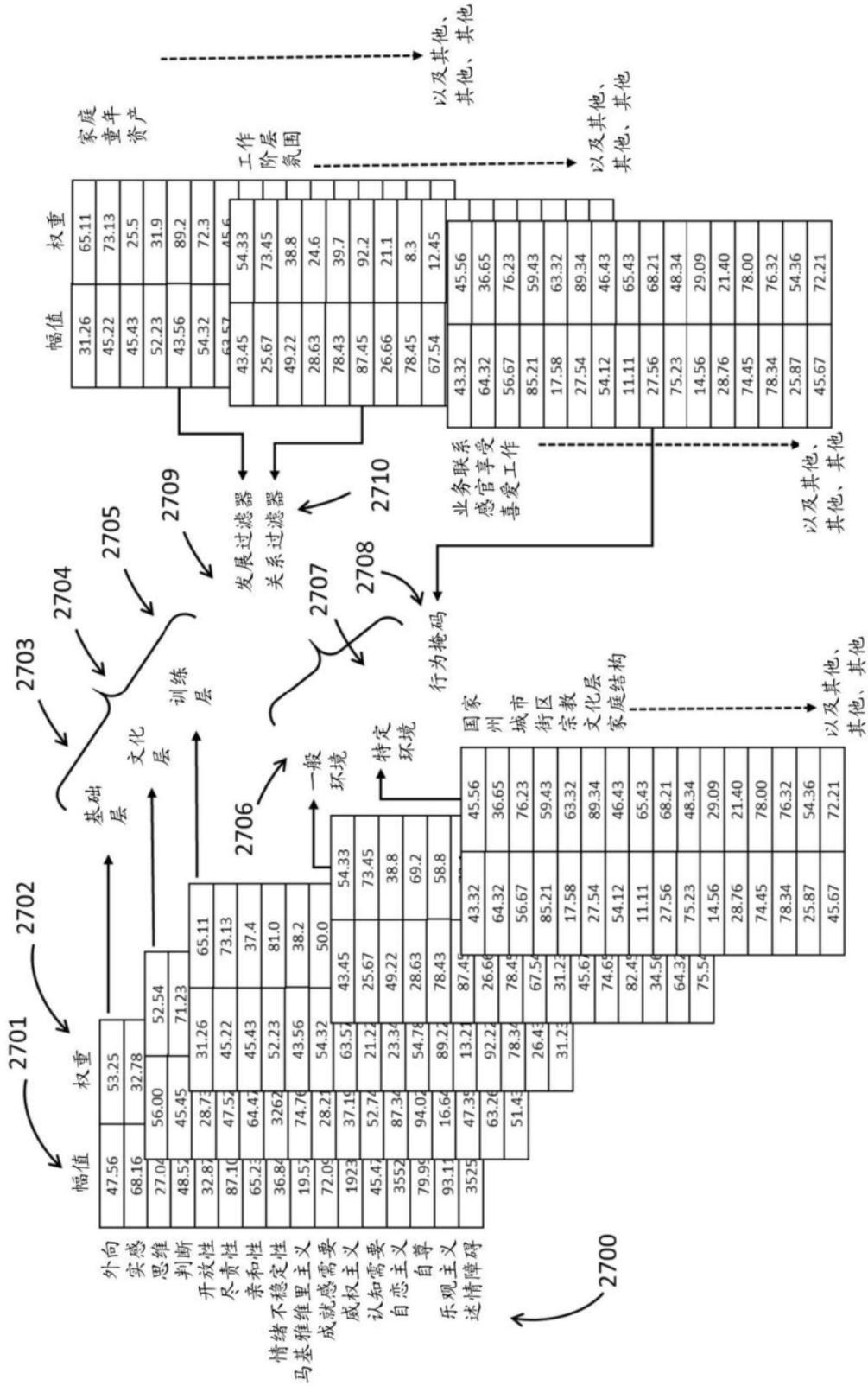


图27

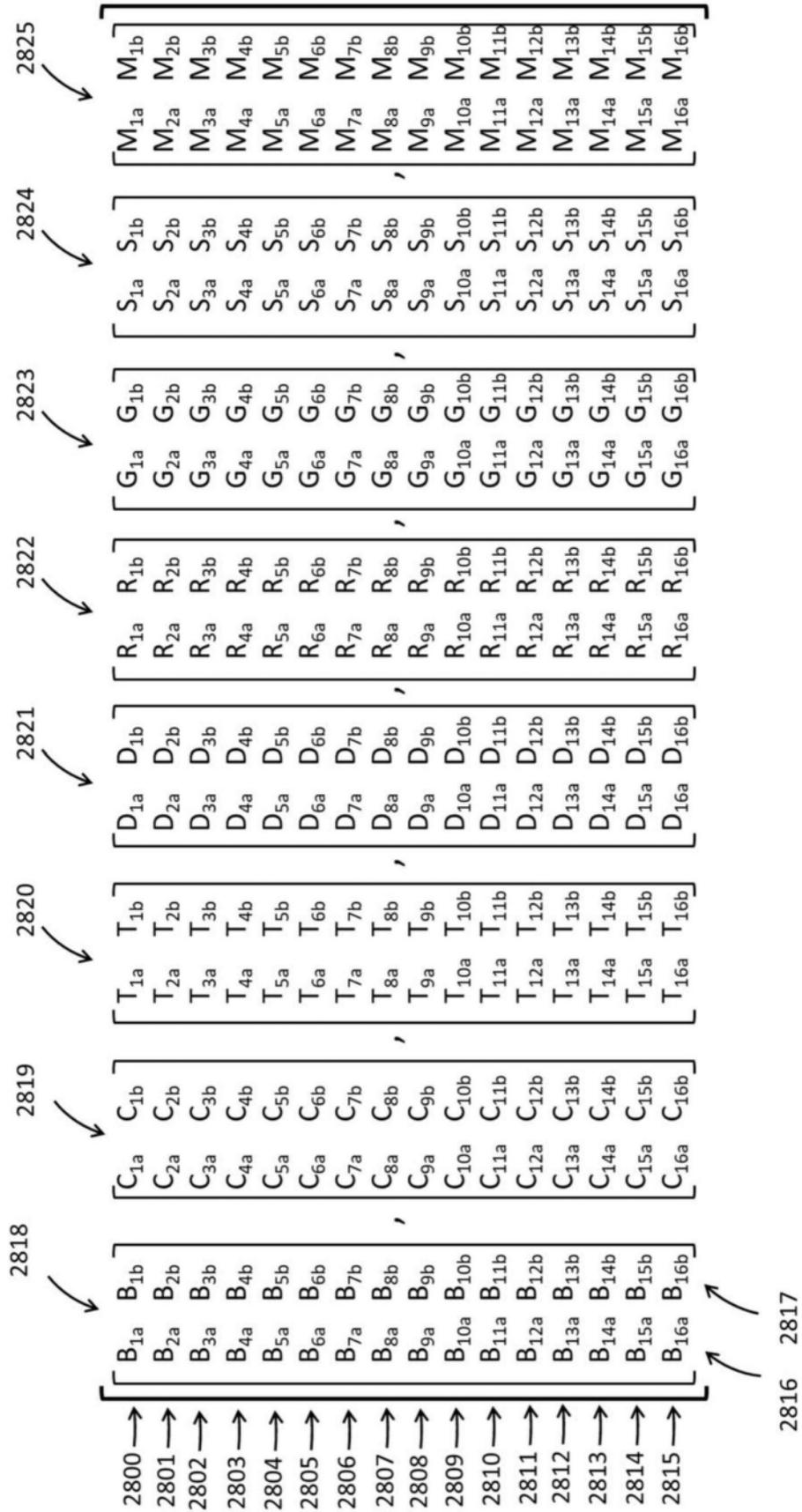


图28

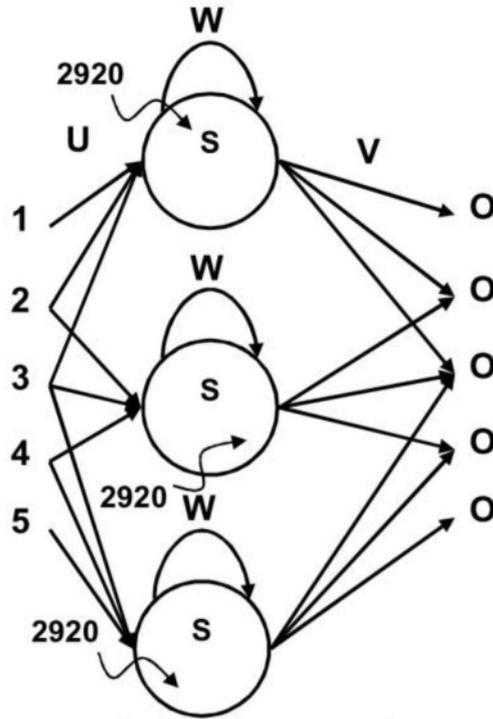


图29A

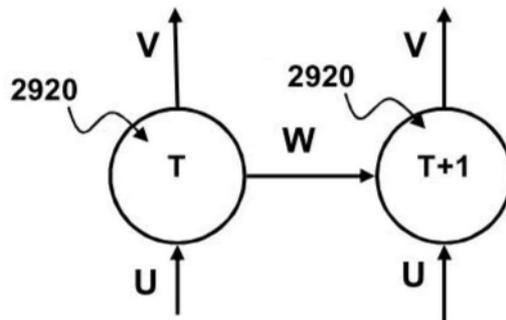


图29B

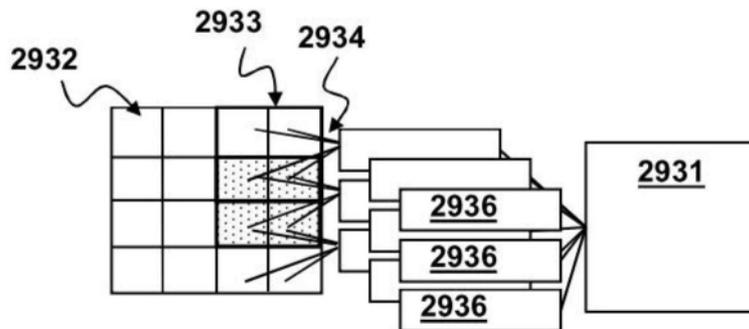


图29C

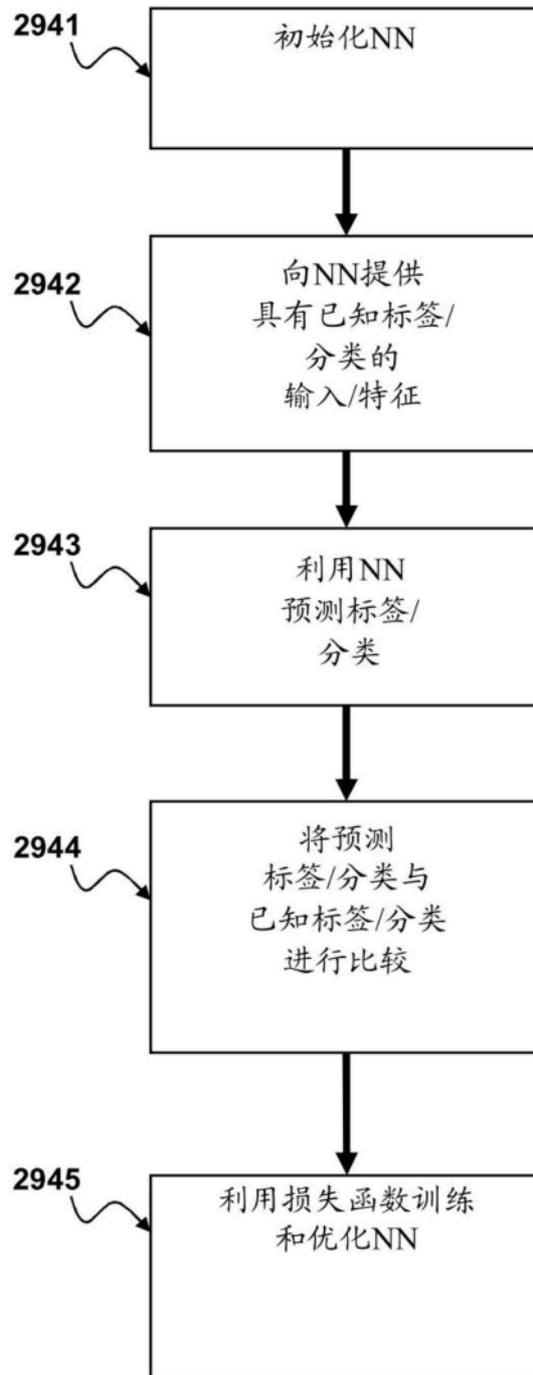


图29D

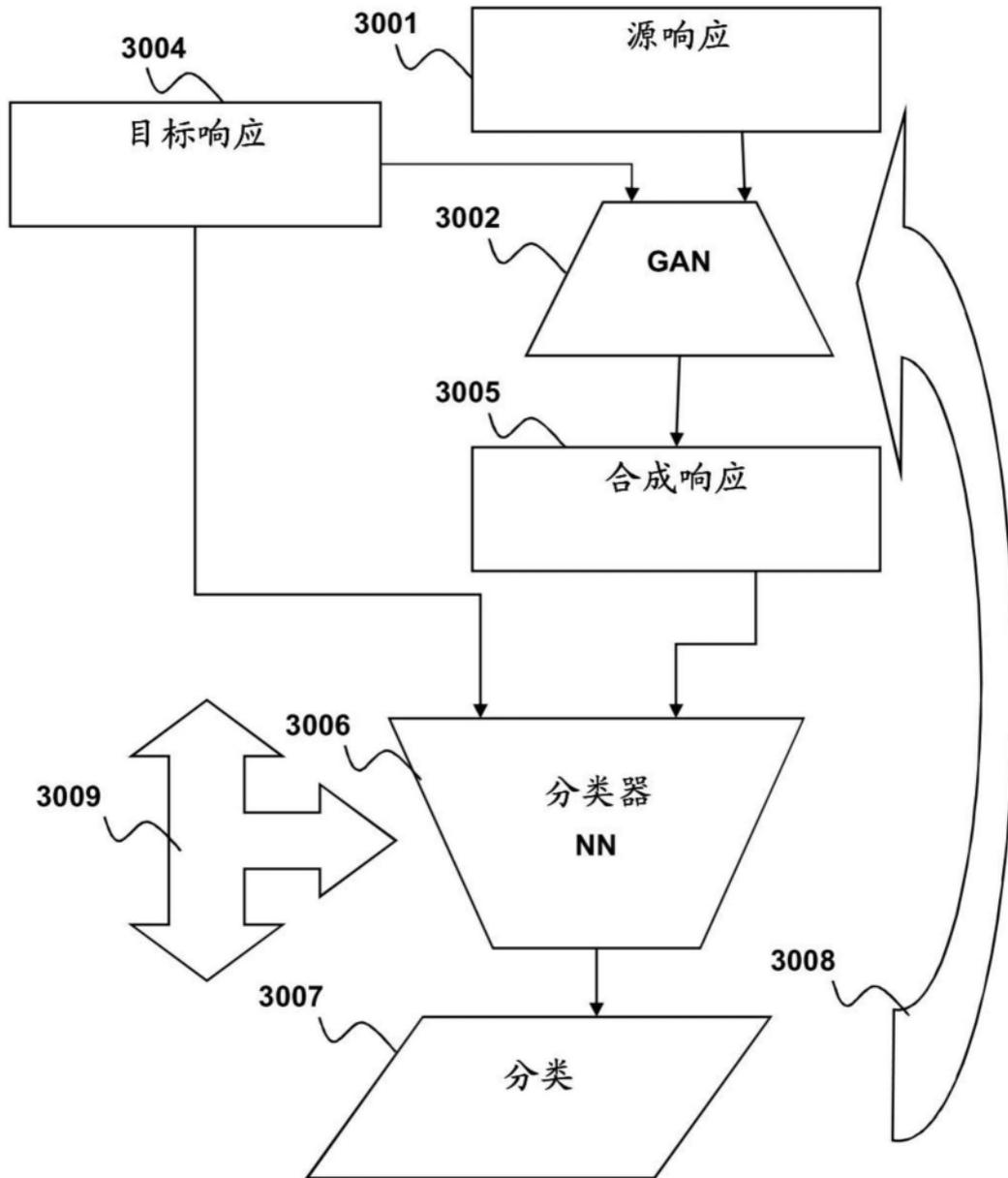


图30

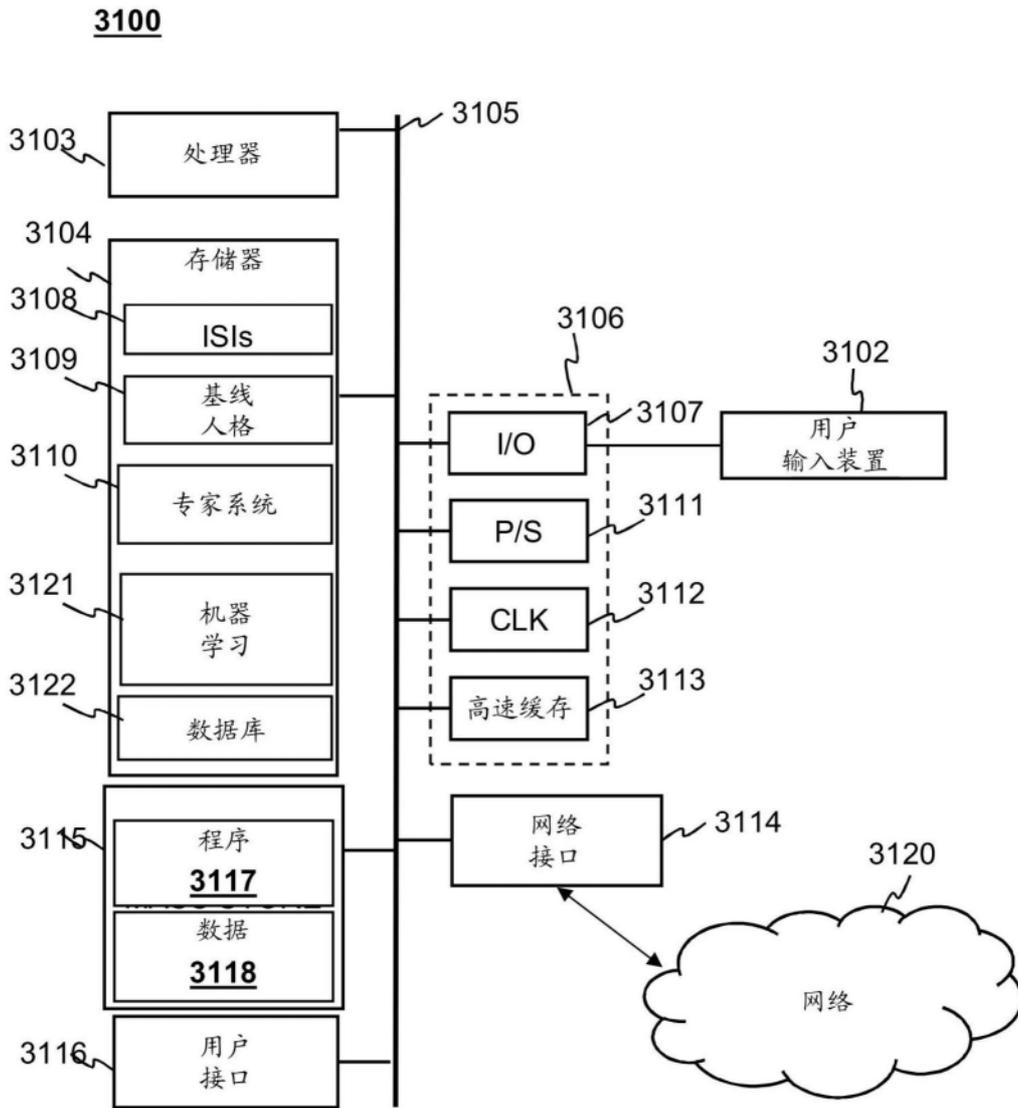


图31